

(19)



Eur päisches Patentamt
European Patent Office
Office européen des brevets



(11)

EP 0 811 688 A2

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:

10.12.1997 Bulletin 1997/50

(21) Application number: 97303918.3

(22) Date of filing: 06.06.1997

(51) Int Cl.⁶: **C12N 15/55**, C12N 9/22,
C07K 16/40, A61K 38/46,
C12Q 1/34, C12Q 1/68
// C12N15/62, (C12N9/22,
C12R1:445)

(84) Designated Contracting States:

BE CH DE DK FR GB IT LI NL

(30) Priority: 06.06.1996 US 19234

01.11.1996 US 29928

01.11.1996 US 29929

(71) Applicants:

- SMITHKLINE BEECHAM CORPORATION
Philadelphia Pennsylvania 19103 (US)
- SMITHKLINE BEECHAM PLC
Brentford, Middlesex TW8 9EP (GB)

(72) Inventors:

- Guth, Sabine, SmithKline Beecham Pharma.
Collegeville, PA 19426 (US)

• Prescott, Catherine D.,
SmithKline Beecham Pharma.
Collegeville, PA 19426 (US)

• Jennings, Joanne,
c/o SmithKline Beecham Pharmac.
Harlow, Essex CM19 5AW (GB)

(74) Representative: Giddings, Peter John, Dr.

SmithKline Beecham plc
Corporate Intellectual Property,
Two New Horizons Court
Brentford, Middlesex TW8 9EP (GB)

(54) **Staphylococcus aureus RNase P**

(57) This invention relates to a novel bacterial ribonucleoprotein complex and the component parts thereof. More specifically, this invention relates to RNase P

isolated from *Staphylococcus aureus* and the use of RNase P or components thereof in screens for the identification of antimicrobial compounds and to the use of such compounds in therapy.

EP 0 811 688 A2

AUG 31 2000

Description**Related Applications:**

5 This application claims benefit of U.S. Provisional Application Number 60/019,234, filed June 6, 1996, U.S. Provisional Application Number 60/029,928, filed November 1, 1996, and U.S. Provisional Application Number 60/029,929, filed November 1, 1996.

Field of the Invention:

10 This invention relates to newly identified polynucleotides, polypeptides encoded by certain of these polynucleotides, molecular complexes of RNAs and polypeptides, the uses of such polynucleotides and polypeptides, as well as the production of such polynucleotides and polypeptides and recombinant host cells transformed with the polynucleotides. The invention relates particularly to such polynucleotides and polypeptides from Staphylococci, especially
15 *S. aureus*. This invention also relates to inhibiting the biosynthesis, assembly or action of such polynucleotides and/or polypeptides and to the use of such inhibitors in therapy.

Background of the Invention:

20 This invention relates to a novel bacterial ribonucleoprotein complex and the component parts thereof. More specifically, this invention relates to RNase P, particularly RNase P from *Staphylococcus aureus*, and the use of RNase P or components thereof in screens for the identification of antimicrobial compounds and to the use of such compounds in therapy.

25 The Staphylococci make up a medically important genera of microbes. They are known to produce two types of disease, invasive and toxigenic. Invasive infections are characterized generally by abscess formation effecting both skin surfaces and deep tissues. *S. aureus* is the second leading cause of bacteremia in cancer patients. Osteomyelitis, septic arthritis, septic thrombophlebitis and acute bacterial endocarditis are also relatively common. There are at least three clinical conditions resulting from the toxigenic properties of Staphylococci. The manifestation of these diseases result from the actions of exotoxins as opposed to tissue invasion and bacteremia. These conditions include: Staphylococcal food poisoning, scalded skin syndrome and toxic shock syndrome.

30 The frequency of *Staphylococcus aureus* infections has risen dramatically in the past 20 years. This has been attributed to the emergence of multiply antibiotic resistant strains and an increasing population of people with weakened immune systems. It is no longer uncommon to isolate *Staphylococcus aureus* strains which are resistant to some or all of the standard antibiotics. This has created a demand for both new anti-microbial agents and diagnostic tests for this organism.

35 While certain Staphylococcal proteins associated with pathogenicity have been identified, e.g., coagulase, hemolysins, leucocidins and exo- and enterotoxins, additional targets are always useful because it is appreciated that the target of an antimicrobial screen can often bias the outcome. Thus, new targets allow for the discovery of new classes of antimicrobials.

Brief Description of the Invention:

40 This invention provides a novel ribonucleoprotein complex, particularly such complex from *Staphylococcus aureus*, and the separately isolated RNA and protein components thereof.

45 In accordance with another aspect of the present invention, there are provided polynucleotides (DNA or RNA) which encode the protein and RNA components of such a complex.

In particular the invention provides polynucleotides having the DNA sequences given herein.

50 The invention also relates to novel oligonucleotides derived from the sequences given herein which can act, for example, as antisense inhibitors of the expression of the RNA or protein components. The oligonucleotides or fragments or derivatives thereof can also be used to directly inhibit catalytic activity or indirectly inhibit activity by interference with RNA protein complex formation. The protein and the RNA components, either separately or in a complex, are also useful as targets in screens designed to identify antimicrobial compounds.

55 It is an object of the invention to provide polypeptides that have been identified as novel RNaseP polypeptides by homology between the amino acid sequence set out in Table 1 [SEQ ID NO: 2] and a known amino acid sequence or sequences of other proteins, such as *B. subtilis* RNase P protein.

It is a further object of the invention to provide polynucleotides that encode RNase P polypeptides, particularly polynucleotides that encode the polypeptide herein designated RNaseP, as well as polynucleotides that are transcribed into RNase P RNA, particularly catalytic RNA.

In a particularly preferred embodiment of the invention the polynucleotide comprises a region encoding RNaseP polypeptides comprising the sequence set out in Table 1 [SEQ ID NO:1] which includes a full length gene, or a variant thereof.

In another particularly preferred embodiment of the invention there is a novel RNaseP protein from *Staphylococcus aureus* comprising the amino acid sequence of Table 1.[SEQ ID NO:2], or a variant thereof.

In accordance with another aspect of the invention there is provided an isolated nucleic acid molecule encoding a mature polypeptide expressible by the *Staphylococcus aureus* WCUH 29 strain contained in the deposited strain.

A further aspect of the invention there are provided isolated nucleic acid molecules encoding RNaseP, particularly *Staphylococcus aureus* RNaseP, including mRNAs, cDNAs, genomic DNAs and catalytic RNAs. Further embodiments of the invention include biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

In accordance with another aspect of the invention, there is provided the use of a polynucleotide of the invention for therapeutic or prophylactic purposes, in particular genetic immunization. Among the particularly preferred embodiments of the invention are naturally occurring allelic variants of RNaseP and polypeptides encoded thereby.

Another aspect of the invention there are provided novel polypeptides of *Staphylococcus aureus* referred to herein as RNaseP as well as biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

Among the particularly preferred embodiments of the invention are variants of RNaseP polypeptide encoded by naturally occurring alleles of the RNaseP gene.

In a preferred embodiment of the invention there are provided methods for producing the aforementioned RNaseP polypeptides.

In accordance with yet another aspect of the invention, there are provided inhibitors to such polypeptides, useful as antibacterial agents, including, for example, antibodies.

In accordance with certain preferred embodiments of the invention, there are provided products, compositions and methods for assessing RNaseP expression, treating disease, for example, disease, such as, infections of the upper respiratory tract (e.g., otitis media, bacterial tracheitis, acute epiglottitis, thyroiditis), lower respiratory (e.g., empyema, lung abscess), cardiac (e.g., infective endocarditis), gastrointestinal (e.g., secretory diarrhoea, splenic abscess, retro-peritoneal abscess), CNS (e.g., cerebral abscess), eye (e.g., blepharitis, conjunctivitis, keratitis, endophthalmitis, preseptal and orbital cellulitis, dacryocystitis), kidney and urinary tract (e.g., epididymitis, intrarenal and perinephric abscess, toxic shock syndrome), skin (e.g., impetigo, folliculitis, cutaneous abscesses, cellulitis, wound infection, bacterial myositis) bone and joint (e.g., septic arthritis, osteomyelitis), assaying genetic variation, and administering a RNaseP polypeptide or polynucleotide to an organism to raise an immunological response against a bacteria, especially a *Staphylococcus aureus* bacteria.

In accordance with certain preferred embodiments of this and other aspects of the invention there are provided polynucleotides that hybridize to RNaseP polynucleotide sequences, particularly under stringent conditions.

In certain preferred embodiments of the invention there are provided antibodies against RNaseP polypeptides.

In other embodiments of the invention there are provided methods for identifying compounds which bind to or otherwise interact with and inhibit or activate an activity of a polypeptide or polynucleotide of the invention comprising: contacting a polypeptide or polynucleotide of the invention with a compound to be screened under conditions to permit binding to or other interaction between the compound and the polypeptide or polynucleotide to assess the binding to or other interaction with the compound, such binding or interaction being associated with a second component capable of providing a detectable signal in response to the binding or interaction of the polypeptide or polynucleotide with the compound; and determining whether the compound binds to or otherwise interacts with and activates or inhibits an activity of the polypeptide or polynucleotide by detecting the presence or absence of a signal generated from the binding or interaction of the compound with the polypeptide or polynucleotide.

In accordance with yet another aspect of the invention, there are provided RNaseP agonists and antagonists, preferably bacteriostatic or bacteriocidal agonists and antagonists.

In a further aspect of the invention there are provided compositions comprising a RNaseP polynucleotide or a RNaseP polypeptide for administration to a cell or to a multicellular organism.

Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following descriptions and from reading the other parts of the present disclosure.

Detailed Description of the Figures:

Figure 1 illustrates the secondary structural modeling of RNase P RNAs based upon published sequence information.

Figure 2 illustrates the sequence alignment of the Staphylococcal RNase P protein component of the invention

compared to other protein components reported in the literature.

Figure 3 illustrates the Staphylococcal DNA encoding a region of the RNase P gene [SEQ ID NO: 14] encoding the RNA component used to clone the intact gene compared with the DNA sequence encoding the gene for the RNase P RNA component from *B. subtilis*.

Figure 4 illustrates the predicted secondary structure for the RNA component of the RNase P of this invention compared to the RNA component from *B. subtilis*.

Figure 5 illustrates the amino acid sequence of the protein component of the *S. aureus* RNase P and the sequence of a DNA encoding same [SEQ ID NOS: 1 and 2].

Figure 6 illustrates a schematic of a whole cell rescue assay to identify drug/RNA interaction.

Figure 7 illustrates a schematic of a screen to identify compounds that disrupt RNA/protein interactions.

Figure 8 illustrates an example of a minimal RNase P substrate embodiment [SEQ ID NO: 4].

Figure 9 shows a schematic diagram of a specific embodiment of the first amplification step of an RNase P gene.

Figure 10 shows a schematic diagram of a specific embodiment of the second amplification step of an RNase P gene.

Glossary:

The following definitions are provided to facilitate understanding of certain terms used frequently herein. Certain other definitions are provided elsewhere herein.

"Host cell" is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous polynucleotide sequence.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Altschul, S.F. et al., *J. Molec. Biol.* 215: 403-410 (1990). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., *J. Mol. Biol.* 215: 403-410 (1990)). As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence of SEQ ID NO: 1 it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence of SEQ ID NO: 1. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. Analogously, by a polypeptide having an amino acid sequence having at least, for example, 95% identity to a reference amino acid sequence of SEQ ID NO:2 is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid of SEQ ID NO: 2. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

"Isolated" means altered "by the hand of man" from its natural state, i.e., if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in

a living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide(s)" generally refers to any polyribonucleotide or polydeoxynucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotide(s)" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions or single-, double- and triple-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded, or triple-stranded regions, or a mixture of single- and double-stranded regions. In addition, "polynucleotide" as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide. As used herein, the term "polynucleotide(s)" also includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotide(s)" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term "polynucleotide(s)" as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including, for example, simple and complex cells. "Polynucleotide(s)" also embraces short polynucleotides often referred to as oligonucleotide(s).

"Polypeptide(s)" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds. "Polypeptide(s)" refers to both short chains, commonly referred to as peptides, oligopeptides and oligomers and to longer chains generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene encoded amino acids. "Polypeptide(s)" include those modified either by natural processes, such as processing and other post-translational modifications, but also by chemical modification techniques. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature, and they are well known to those of skill in the art. It will be appreciated that the same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains, and the amino or carboxyl termini. Modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins, such as arginylation, and ubiquitination. See, for instance, *PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993) and Wold, F., *Posttranslational Protein Modifications: Perspectives and Prospects*, pgs. 1-12 in *POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS*, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter et al., *Meth. Enzymol.* 182:626-646 (1990) and Rattan et al., *Protein Synthesis: Posttranslational Modifications and Aging*, Ann. N.Y. Acad. Sci. 663: 48-62 (1992). Polypeptides may be branched or cyclic, with or without branching. Cyclic, branched and branched circular polypeptides may result from post-translational natural processes and may be made by entirely synthetic methods, as well.

"Variant(s)" as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques, by direct synthesis, and by other recombinant methods known to skilled artisans.

Detail d Description of th Invention:

The ribonucleoprotein, RNase P, plays a key role in the biosynthesis of transfer RNA (tRNA), itself a key intermediate in protein biosynthesis. RNase P functions to process precursor RNAs into mature tRNAs by endoribonucleolytic action. The complex in prokaryotes is composed of two subunits: a catalytic RNA and protein co-factor. Recent reviews of certain RNase P molecules exist. See for example, L.A. Kirsebom, Molecular Microbiology 17(3), 411-420 (1995) or N. R. Pace and J. W. Brown, J. Bacteriol. 177 (8), 1919-1928(1995).

The invention relates to novel RNaseP polypeptides and polynucleotides as described in greater detail below. In particular, the invention relates to polypeptides and polynucleotides of a novel RNaseP of *Staphylococcus aureus*, which is related by amino acid sequence homology to RNase P polypeptide set forth in Figures 2 and 5 and SEQ ID NO:2. The invention relates especially to RNaseP having the nucleotide and amino acid sequences set out in Table 1, SEQ ID NO: 1 and SEQ ID NO: 2 respectively, and to the RNaseP nucleotide sequences of the DNA in the deposited strain and amino acid sequences encoded thereby. The invention also relates to the RNase P RNA component, particularly in its catalytic for, and sequences from which such component is transcribed.

RNase P RNA Component:

Phylogenetic comparisons readily allow secondary structure modeling and the identification of a minimum consensus structure (Figure 1). Data concerning the RNase P RNA structure are available in the RNase P database on the w.w.w. (<http://jwbrown.mbio.ncsu.edu/RNaseP/home.html>). A polynucleotide of the invention from which the RNA component is transcribed is set forth in Figure 3 [SEQ ID NO: 14].

In general, few nucleotides are conserved but compensatory base changes in hydrogen bonded regions show that the overall structure is preserved throughout eubacteria. Universal conservation of primary sequences (*E. coli* : 61-74, 353-360) together with other conserved or quasi-conserved nucleotides implicate functional importance, the significance of which remain unknown. To date all RNase P RNA molecules can be folded to fit a consensus 'cage-like' structure, beyond this domain there is no convincing structural similarity between prokaryotic and eukaryotic RNase P RNAs.

RNase P Protein Component:

The precise functional role of the protein remains unknown. While it is appreciated that *in vitro* the experimental conditions can be established such that the protein component is not necessary for catalytic activity, *in vivo*, the protein components appears to be required. A novel RNase P protein component from *S. aureus* has been identified and is characterized by the amino acid sequence given in Figure 2 [SEQ ID NO:2] in which the *S. aureus* (Sau) sequence is aligned with sequences from other microorganisms.

The full length sequence encoding the intact RNase P protein component can be obtained by probing a genomic library by for example in situ colony hybridization detailed in Maniatis *et al.* (*infra*) using a probe(s) generated based on the sequences given in Figure 5 [SEQ ID NOS: 1 and 2].

TABLE 1
RNaseP Polynucleotide and Polypeptide Sequences

(A) Sequences from *Staphylococcus aureus* RNaseP polynucleotide sequence [SEQ ID NO:1].

5' -ATG TTA TTG GAA AAA GCT TAC CGA ATT AAA AAG AAT GCA GAT
TTT CAG AGA ATA TAT AAA AAA GGT CAT TCT GTA GCC AAC AGA CAA
TTT GTT GTA TAC ACT TGT AAT AAT AAA GAA ATA GAC CAT TTT CGC
TTA GGT ATT AGT GTT TCT AAA AAA CTA GGT AAT GCA GTG TTA AGA
AAC AAG ATT AAA AGA GCA ATA CGT GAA AAT TTC AAA GTA CAT AAG
TCG CAT ATA TTG GCC AAA GAT ATT ATT GTA ATA GCA AGA CAG CCA
GCT AAA GAT ATG ACG ACT TTA CAA ATA CAG AAT AGT CTT GAG CAC
GTA CTT AAA ATT GCC AAA GTT TTT AAT AAA AAG ATT AAG TAA-3'

(B) RNaseP polypeptide sequence deduced from the polynucleotide sequence in this table [SEQ ID NO:2].

NH₂- MLLEK VYRIK KNADF GRIYK KGHSV ANRQF VVYTC NNKEI DHFRL
GISVS KKLGN AVLRL KIKRA IRENF KVHKS HILAK DIIVI ARQPA KDMTT
LQIQN SLEHV LKIAK VFNKK IK-COOH

(C) Polynucleotide sequence embodiments [SEQ ID NO:1].

X-(R₁)_n-ATG TTA TTG GAA AAA GCT TAC CGA ATT AAA AAG AAT GCA
GAT TTT CAG AGA ATA TAT AAA AAA GGT CAT TCT GTA GCC AAC AGA
CAA TTT GTT GTA TAC ACT TGT AAT AAT AAA GAA ATA GAC CAT TTT
CGC TTA GGT ATT AGT GTT TCT AAA AAA CTA GGT AAT GCA GTG TTA
AGA AAC AAG ATT AAA AGA GCA ATA CGT GAA AAT TTC AAA GTA CAT
AAG TCG CAT ATA TTG GCC AAA GAT ATT ATT GTA ATA GCA AGA CAG
CCA GCT AAA GAT ATG ACG ACT TTA CAA ATA CAG AAT AGT CTT GAG
CAC GTA CTT AAA ATT GCC AAA GTT TTT AAT AAA AAG ATT AAG TAA-
(R₂)_n-Y

(D) Polypeptide sequence embodiments [SEQ ID NO:2].

X-(R₁)_n-MLLEK VYRIK KNADF GRIYK KGHSV ANRQF VVYTC NNKEI DHFRL
GISVS KKLGN AVLRL KIKRA IRENF KVHKS HILAK DIIVI ARQPA KDMTT
LQIQN SLEHV LKIAK VFNKK IK-(R₂)_n-Y

(E) Sequences from *Staphylococcus aureus* RNaseP RNA gene [SEQ ID NO:3].

5' -GTT CTGAT ATTT CGGGT AAT CG CT AT ATT AT AT AGAGGAAAGT CCATG CT CCACAGT CTGAGATGCATT
GTAGT GTT CGT G CTT GATGAAA CAAT AAAT CAAGGCATT AATTT GAOCGCAATGAAAT AT CCT AAGT CT
TT CGAT ATGGAT AGAGT AATTT GAAAGTG CCA CAGTGA CGT AGCTTTT AT AGAAAT AT AAAAGGTGGAA
CGCGGT AAA CCCCT CGAGT GAG CAAT CCAAATTT GGT AGGAGCA CTTGTTT AACGGAATT CAA CGT AT
AAA CGAGACA CATT CG CGAAATGAAGTGGTGTACGACAGATGGTTAT CACCTGAGT ACCAGTGTGA
CTAGTGCA CGT GATGAGTACGATGGAACAGAACATGGCTT AT AGAAAT AT CACTACTA
G-3'

(F) Polynucleotide sequence embodiments [SEQ ID NO:3].

X-(R₁)_n-
GTT CTGAT ATTT CGGGT AAT CG CT AT ATT AT AT AGAGGAAAGT CCATG CT CCACAGT CTGAGATGCATT
GTAGT GTT CGT G CTT GATGAAA CAAT AAAT CAAGGCATT AATTT GAOCGCAATGAAAT AT CCT AAGT CT
TT CGAT ATGGAT AGAGT AATTT GAAAGTG CCA CAGTGA CGT AGCTTTT AT AGAAAT AT AAAAGGTGGAA
CGCGGT AAA CCCCT CGAGT GAG CAAT CCAAATTT GGT AGGAGCA CTTGTTT AACGGAATT CAA CGT AT

AAACGAGACACACTT CGCGAAATGAAGTGGTGTACGACAGATGGTTATCACCTGAGTACCAAGTGTGA
 CTAGTGCACTGATGAGTACGATGGAAAGAAATCATGGCTTATAGAAATATCACTACTA
 5 G-(R₂)_n-Y

Polypeptides of the invention:

The polypeptides of the invention include the polypeptide of Table 1 [SEQ ID NO:2] (in particular the mature polypeptide) as well as polypeptides and fragments, particularly those which have the biological activity of RNaseP, and also those which have at least 70% identity to the polypeptide of Table 1 [SEQ ID NO:2] or the relevant portion, preferably at least 80% identity to the polypeptide of Table 1 [SEQ ID NO:2], and more preferably at least 90% similarity (more preferably at least 90% identity) to the polypeptide of Table 1 [SEQ ID NO:2] and still more preferably at least 95% similarity (still more preferably at least 95% identity) to the polypeptide of Table 1 [SEQ ID NO:2] and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.

The invention also includes polypeptides of the formula set forth in Table 1(D) [SEQ ID NO:2] wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal, R₁ and R₂ is any amino acid residue, and n is an integer between 1 and 1000. Any stretch of amino acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

A fragment is a variant polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of the aforementioned polypeptides. As with RNaseP polypeptides fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region, a single larger polypeptide.

Preferred fragments include, for example, truncation polypeptides having a portion of the amino acid sequence of Table 1 [SEQ ID NO:2], or of variants thereof, such as a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus. Degradation forms of the polypeptides of the invention in a host cell, particularly a *Staphylococcus aureus*, are also preferred. Further preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

Also preferred are biologically active fragments which are those fragments that mediate activities of RNaseP, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those fragments that are antigenic or immunogenic in an animal, especially in a human. Particularly preferred are fragments comprising receptors or domains of enzymes that confer a function essential for viability of *Staphylococcus aureus* or the ability to initiate, or maintain cause disease in an individual, particularly a human.

Variants that are fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, these variants may be employed as intermediates for producing the full-length polypeptides of the invention.

Polynucleotides of the invention:

Another aspect of the invention relates to isolated polynucleotides, including the full length gene, that encode the RNaseP polypeptide having a deduced amino acid sequence of Table 1 [SEQ ID NO:2] and polynucleotides closely related thereto and variants thereof.

Using the information provided herein, such as a polynucleotide sequence set out in SEQ ID NOS: 1, 3, 4 and 14, a polynucleotide of the invention encoding RNaseP polypeptide or RNA (such as that transcribed from SEQ ID NO:3) may be obtained using standard cloning and screening methods, such as those for cloning and sequencing chromosomal DNA fragments from bacteria using *Staphylococcus aureus* WCUH 29 cells as starting material, followed by obtaining a full-length clone. For example, to obtain a polynucleotide sequence of the invention, such as a sequence given in SEQ ID NOS: 1, 3, 4 and 14, typically a library of clones of chromosomal DNA of *Staphylococcus aureus* WCUH 29 in *E. coli* or some other suitable host is probed with a radiolabeled oligonucleotide, preferably a 17-mer or longer, derived from a partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using stringent conditions. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence. Conveniently, such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook et al., *MOLECULAR CLONING, A LAB-*

ORATORY MANUAL, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). (see in particular Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70). Illustrative of the invention, the polynucleotides set out in Table I [SEQ ID NO:1, 4 and 5] were discovered in a DNA library derived from *Staphylococcus aureus* WCUH 29.

Certain DNA sequences set out in Table 1 [SEQ ID NO:1] contains an open reading frame encoding a protein having about the number of amino acid residues set forth in Table 1 [SEQ ID NO:2] with a deduced molecular weight that can be calculated using amino acid residue molecular weight values well known in the art. The polynucleotide of SEQ ID NO: 1, between nucleotide number 1 through number 351 encodes the polypeptide of SEQ ID NO:2. The stop codon begins at nucleotide number 352 of SEQ ID NO: 1.

RNaseP of the invention is structurally related to other proteins of the RNase P family, as shown by the results of sequencing the DNA encoding RNaseP of the deposited strain. The protein exhibits greatest homology to *B. subtilis* protein among known proteins. RNaseP of Table I [SEQ ID NO:2] has significant identity and similarity over its entire length with the amino acid sequence of *B. subtilis* RNase P polypeptide.

The invention provides a polynucleotide sequence identical over its entire length to the coding sequence in Table I [SEQ ID NO:1]. Also provided by the invention is the coding sequence for the mature polypeptide or a fragment thereof, by itself as well as the coding sequence for the mature polypeptide or a fragment in reading frame with other coding sequence, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro-protein sequence. The polynucleotide may also contain non-coding sequences, including for example, but not limited to non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence which encode additional amino acids. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain embodiments of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc. Natl. Acad. Sci., USA* 86: 821-824 (1989), or an HA tag (Wilson *et al.*, *Cell* 37: 767 (1984). Polynucleotides of the invention also include, but are not limited to, polynucleotides comprising a structural gene and its naturally associated sequences that control gene expression.

A preferred embodiment of the invention is a polynucleotide of comprising nucleotide 1 to 351 or 354 set forth in SEQ ID NO: 1 of Table 1 which encode the RNaseP polypeptide.

The invention also includes polynucleotides of the formula set forth in Table 1 (C)[SEQ ID NO:1 and (F)[SEQ ID NO:3] wherein, at the 5' end of the molecule, X is hydrogen, and at the 3' end of the molecule, Y is hydrogen or a metal, R₁ and R₂ is any nucleic acid residue, and n is an integer between 1 and 3000. Any stretch of nucleic acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer. A preferred embodiment for the sequence set forth in Table 1(F) [SEQ ID NO:3] has R₁ or R₂ being between 1 and 10 or 1 and 20, and especially being 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10. The invention also provides RNA transcribed from such polynucleotides, particularly catalytic RNAs.

The term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides that include a sequence encoding a polypeptide of the invention, particularly a bacterial polypeptide and more particularly a polypeptide of the *Staphylococcus aureus* RNaseP having the amino acid sequence set out in Table 1 [SEQ ID NO:2]. The term also encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, interrupted by integrated phage or an insertion sequence or editing) together with additional regions, that also may contain coding and/or non-coding sequences.

The invention further relates to variants of the polynucleotides described herein that encode for variants of the polypeptide having the deduced amino acid sequence of Table 1 [SEQ ID NO:2]. Variants that are fragments of the polynucleotides of the invention may be used to synthesize full-length polynucleotides of the invention.

In addition to the standard A, G, C, T/U representations for nucleic acid bases, the term "N" is also used. "N" means that any of the four DNA or RNA bases may appear at such a designated position in the DNA or RNA sequence, except that, in preferred embodiments, N can not be a base that when taken in combination with adjacent nucleotide positions, when read in the correct reading frame, would have the effect of generating a premature termination codon in such reading frame.

Further particularly preferred embodiments are polynucleotides encoding RNaseP variants, that have the amino acid sequence of RNaseP polypeptide of Table 1 [SEQ ID NO:2] in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, that do not alter the properties and activities of RNaseP.

Further preferred embodiments of the invention are polynucleotides that are at least 50%, 60% or 70% identical over their entire length to a polynucleotide encoding RNaseP polypeptide having an amino acid sequence set out in Table 1 [SEQ ID NO:2], and polynucleotides that are complementary to such polynucleotides. Alternatively, most highly preferred are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding RNaseP polypeptide of the deposited strain and polynucleotides complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length to the same are particularly preferred, and among these

particularly preferred polynucleotides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the more preferred.

Preferred embodiments of the invention are polynucleotides that are at least 50%, 60% or 70% identical over their entire length to an RNaseP polynucleotide having a nucleotide sequence set out in SEQ ID NO:3, 4 or 14, and polynucleotides that are complementary to such polynucleotides. Alternatively, most highly preferred are polynucleotides that comprise a region that is at least 80% identical over its entire length to an RNaseP polynucleotide of the deposited strain and polynucleotides complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length to the same are particularly preferred, and among these particularly preferred polynucleotides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the more preferred. It is especially preferred that these polynucleotides be RNAs, especially catalytic RNAs.

Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptide encoded by the DNA of Table 1 [SEQ ID NO: 1] or as the RNase P RNA component transcribed by the DNA of SEQ ID NO:3, 4 or 14.

The invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the invention especially relates to polynucleotides that hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the terms "stringent conditions" and "stringent hybridization conditions" mean hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at about 65°C. Hybridization and wash conditions are well known and exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), particularly Chapter 11 therein.

The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set forth in SEQ ID NO: 1 or SEQ ID NO:3 or SEQ ID NO:4 or SEQ ID NO:14 under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence set forth in SEQ ID NO: 1 or SEQ ID NO:3 or SEQ ID NO:4 or SEQ ID NO:14 respectively or a fragment thereof; and isolating said DNA sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers described elsewhere herein.

As discussed herein regarding polynucleotide assays of the invention, for instance, polynucleotides of the invention as discussed above, may be used as a hybridization probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding RNaseP and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to the RNaseP gene. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 30 bases and may have at least 50 bases. Particularly preferred probes will have at least 30 bases and will have 50 bases or less.

Polynucleotides of the invention that are oligonucleotides derived from the sequences of SEQ ID NOS:1 and/or 2 and/or 3 and/or 4 and/or 14 may be used in the processes herein as described, but preferably for PCR, to determine whether or not the polynucleotides identified herein in whole or in part are transcribed in bacteria in infected tissue. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The invention also provides polynucleotides that may encode a polypeptide that is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may allow protein transport, may lengthen or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in vivo*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

In sum, a polynucleotide of the invention may encode a mature protein, a mature protein plus a leader sequence (which may be referred to as a preprotein), a precursor of a mature protein having one or more prosequences that are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more prosequences, which generally are removed during processing steps that produce active and mature forms of the polypeptide.

Cloning of *S. aureus* RNase P RNA structural gene:

A partial homolog to *B. subtilis* RNase P RNA was identified in a proprietary *S. aureus* database comprising sequences from randomly sequenced *S. aureus* DNA library in an *E. coli* host. This homolog (A33202: *B. subtilis* RNA nucleotides 311-460) is shown in Figure 3 [SEQ ID NO:14].

A PCR primer based on these data was designed to the 3' end of the gene (primer: 5'-CGC GAA GTG TGT CTC GTT TAT ACG-3') [SEQ ID NO:5] and a second based on a universally conserved sequence within the 5' domain (5'-GAG GAA AGT CCA TGC TC-3') [SEQ ID NO:6] (available from the RNase P database w.w.w.) permitted recovery of approximately 90% of the gene. The complete structural gene was amplified using a degenerate primer (5'-C/TGA-TATTTCTG/TGA/GTAAT/C-3') [SEQ ID NO:13] that would allow the RNA product to form the predicted helices P1 and P2 (Figure 1). The structural gene has been cloned behind a T7 promoter, sequenced and shown to be highly related to *B. subtilis* homolog (Figure 2). The precise *S. aureus* genomic sequence encoding (transcribing) helices P1 and P2 may be determined by skilled artisans using methods and compounds of the invention, such as the degenerate primer described above [SEQ ID NO: 13].

The predicted secondary structure is shown in Figure 4 and based on data available for *B. subtilis*, reveals a limited number of compensatory base changes within helices (e.g. P1, P2, P8, P12), an apparent shortening of helices P3 and P9 and an extension of helix P18.

Further Description and Definitions:

The coding region of the RNaseP gene may be isolated, for example, by screening using a deposit containing a *Staphylococcus aureus* WCUH 29 strain which has been deposited with the National Collections of Industrial and Marine Bacteria Ltd. (herein "NCIMB"), 23 St. Machar Drive, Aberdeen AB2 1RY, Scotland on 11 September 1995 and assigned NCIMB Deposit No. 40771. It was referred to as *Staphylococcus aureus* WCUH29 on deposit. The *Staphylococcus aureus* strain deposit is referred to herein as "the deposited strain" or as "the DNA of the deposited strain."

The deposited strain contains the full length RNaseP gene. The sequence of the polynucleotides contained in the deposited strain, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

The deposit of the deposited strain has been made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for Purposes of Patent Procedure. The strain will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. The deposited strain is provided merely as convenience to those of skill in the art and is not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. §112.

A license may be required to make, use or sell the deposited strain, and compounds derived therefrom, and no such license is hereby granted.

The nucleotide sequences disclosed herein can also be obtained by synthetic chemical techniques known in the art or can be obtained from *S. aureus* WCUH 29 by probing a DNA preparation with probes constructed from the particular sequences disclosed herein. Alternatively, oligonucleotides derived from a disclosed sequence can act as PCR primers in a process of PCR-based cloning of the sequence from a bacterial genomic source. It is recognized that such sequences will also have utility in diagnosis of the type of infection the pathogen has attained.

A polynucleotide of the present invention may be in the form of RNA or in the form of DNA, which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequence which encodes the polypeptide may be identical to the coding sequence shown or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encoding the same polypeptide.

Thus, the term "polynucleotide encoding a polypeptide" encompasses a polynucleotide which includes only coding sequence for the polypeptide as well as a polynucleotide which includes additional coding and/or non-coding sequence.

The present invention therefore includes polynucleotides, wherein the coding sequence for the mature polypeptide may be fused in the same reading frame to a polynucleotide sequence which aids in expression and secretion of a polypeptide from a host cell, for example, a leader sequence which functions as a secretory sequence for controlling transport of a polypeptide from the cell. The polypeptide having a leader sequence is a preprotein and may have the leader sequence cleaved by the host cell to form the mature form of the polypeptide. The polynucleotides may also encode for a proprotein which is the mature protein plus additional 5' amino acid residues. A mature protein having a prosequence is a proprotein and is an inactive form of the protein. Once the prosequence is cleaved an active mature protein remains.

Thus, for example, the polynucleotide of the present invention may encode for a mature protein, or for a protein having a prosequence or for a protein having both a prosequence and a presequence (leader sequence). Further, the amino acid sequences provided herein show a methionine residue at the NH₂-terminus. It is appreciated, however,

that during post-translational modification of the peptide, this residue may be deleted. Accordingly, this invention contemplates the use of both the methionine-containing and the methionineless amino terminal variants of each protein disclosed herein.

The polynucleotides of the present invention may also have the coding sequence fused in frame to a marker sequence at either the 5' or 3' terminus of the gene which allows for purification of the polypeptide of the present invention. The marker sequence may be a hexa-histidine tag supplied by the pQE series of vectors (supplied commercially by Quiagen Inc.) to provide for purification of the polypeptide fused to the marker in the case of a bacterial host. Alternatively the maltose binding protein (MBP) fusion system may be employed. In this system the gene of interest is fused the malE gene encoding the MBP (supplied by New England BioLabs). The fusion product is purified in a one step procedure based on the MBP affinity for maltose. A pre-engineered Xa cleavage site allows for efficient removal of the MBP component from the gene product of interest.

In order to facilitate understanding of the following example certain frequently occurring methods and/or terms will be described.

"Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

"Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 µg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 µl of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on an agarose gel to isolate the desired fragment. Size separation of the cleaved fragments is generally performed using a 1% percent agarose gel.

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T., *et al.*, supra, p. 146). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units to T4 DNA ligase ("ligase") per 0.5 µg of approximately equimolar amounts of the DNA fragments to be ligated.

The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

A "replicon" is any genetic element (e.g., plasmid, chromosome, virus) that functions as an autonomous unit of DNA replication *in vivo*; i.e., capable of replication under its own control.

A "vector" is a replicon, such as a plasmid, phage, or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment.

A "double-stranded DNA molecule" refers to the polymeric form of deoxyribonucleotides (bases adenine, guanine, thymine, or cytosine) in a double-stranded helix, both relaxed and supercoiled. This term refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear DNA molecules (e.g., restriction fragments), viruses, plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the nontranscribed strand of DNA (i.e., the strand having the sequence homologous to the mRNA).

A DNA "coding sequence" or a "nucleotide sequence encoding" a particular protein, is a DNA sequence which is transcribed and translated into a polypeptide when placed under the control of appropriate regulatory sequences.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bound at the 3' terminus by a translation start codon (e.g., ATG) of a coding sequence and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site (conveniently defined by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes. Prokaryotic promoters contain the -10 and -35 consensus sequences.

DNA "control sequences" refers collectively to promoter sequences, ribosome binding sites, polyadenylation signals, transcription termination sequences, upstream regulatory domains, enhancers, and the like, which collectively provide for the expression (i.e., the transcription and translation) of a coding sequence in a host cell.

A control sequence "directs the expression" of a coding sequence in a cell when RNA polymerase will bind the promoter sequence and transcribe the coding sequence into mRNA, which is then translated into the polypeptide encoded by the coding sequence.

A "host cell" is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous DNA sequence.

A cell has been "transformed" by exogenous DNA when such exogenous DNA has been introduced inside the cell membrane. Exogenous DNA may or may not be integrated (covalently linked) into chromosomal DNA making up the genome of the cell. In prokaryotes and yeasts, for example, the exogenous DNA may be maintained on an episomal element, such as a plasmid. With respect to eukaryotic cells, a stably transformed or transfected cell is one in which the exogenous DNA has become integrated into the chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cell containing the exogenous DNA.

A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations.

A "heterologous" region of a DNA construct is an identifiable segment of DNA within or attached to another DNA molecule that is not found in association with the other molecule in nature.

Preparation of the RNase P protein component:

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

In accordance with yet a further aspect of the present invention, there is therefore provided a process for producing the polypeptide of the invention by recombinant techniques by expressing a polynucleotide encoding said polypeptide in a host and recovering the expressed product. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a cosmid, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

Suitable expression vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA. However, any other vector may be used as long as it is replicable and viable in the host.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof or polynucleotides of the invention. Introduction of a polynucleotide into the host cell can be effected by methods described in many standard laboratory manuals, such as Davis et al., *BASIC METHODS IN MOLECULAR BIOLOGY*, (1986) and Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), such as, calcium phosphate transfection, DEAE-dextran mediated transfection, transfection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction and infection.

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, enterococci *E. coli*, streptomyces and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used to produce the polypeptides of the invention. Such vectors include, among others; chromosomal, episomal and virus-derived vectors; e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression system constructs may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides and/or to express a polypeptide in a host may be used for expression in this regard. The appropriate DNA sequence may be inserted into the expression system by any of a variety of well-

known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL*, (*supra*).

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence (s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the *E. coli* *lac* or *trp*, the phage lambda P_L promoter and other promoters known to control expression of genes in eukaryotic or prokaryotic cells or their viruses. The expression vector may also contain a ribosome binding site for translation initiation and/or a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

The gene can be placed under the control of a promoter, ribosome binding site (for bacterial expression) and, optionally, an operator (collectively referred to herein as "control" elements), so that the DNA sequence encoding the desired protein is transcribed into RNA in the host cell transformed by a vector containing this expression construction. The coding sequence may or may not contain a signal peptide or leader sequence. The polypeptides of the present invention can be expressed using, for example, the *E. coli* *lac* promoter or the protein A gene (*spa*) promoter and signal sequence. Leader sequences can be removed by the bacterial host in post-translational processing. See, e.g., U.S. Patent Nos. 4,431,739; 4,425,437; 4,338,397. Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are PKK232-8 and PCM7. Particular named bacterial promoters include *lacI*, *lacZ*, T3, T7, *gpt*, lambda P_R, P_L and *trp*. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In addition to control sequences, it may be desirable to add regulatory sequences which allow for regulation of the expression of the protein sequences relative to the growth of the host cell. Regulatory sequences are known to those of skill in the art, and examples include those which cause the expression of a gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Other types of regulatory elements may also be present in the vector, for example, enhancer sequences.

An expression vector is constructed so that the particular coding sequence is located in the vector with the appropriate regulatory sequences, the positioning and orientation of the coding sequence with respect to the control sequences being such that the coding sequence is transcribed under the "control" of the control sequences (i.e., RNA polymerase which binds to the DNA molecule at the control sequences transcribes the coding sequence). Modification of the coding sequences may be desirable to achieve this end. For example, in some cases it may be necessary to modify the sequence so that it may be attached to the control sequences with the appropriate orientation; i.e., to maintain the reading frame. The control sequences and other regulatory sequences may be ligated to the coding sequence prior to insertion into a vector, such as the cloning vectors described above. Alternatively, the coding sequence can be cloned directly into an expression vector which already contains the control sequences and an appropriate restriction site.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example. Bacterial: pET-3 vectors (Stratagene), pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, PKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pBlueBacII (Invitrogen), pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any

other plasmid or vector may be used as long as they are replicable and viable in the host.

Examples of recombinant DNA vectors for cloning and host cells which they can transform include the bacteriophage λ (*E. coli*), pBR322 (*E. coli*), pACYC177 (*E. coli*), pKT230 (gram-negative bacteria), pGV1106 (gram-negative bacteria), pLAFR1 (gram-negative bacteria), pME290 (non-*E. coli* gram-negative bacteria), pHV14 (*E. coli* and *Bacillus subtilis*), pBD9 (*Bacillus*), pJ61 (*Streptomyces*), pUC6 (*Streptomyces*), Ylp5 (*Saccharomyces*), a baculovirus insect cell system, YCp19 (*Saccharomyces*). See, generally, "DNA Cloning": Vols. I & II, Glover *et al.* ed. IRL Press Oxford (1985) (1987) and: T. Maniatis *et al.* ("Molecular Cloning" Cold Spring Harbor Laboratory (1982).

In some cases, it may be desirable to add sequences which cause the secretion of the polypeptide from the host organism, with subsequent cleavage of the secretory signal.

Polypeptides can be expressed in host cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well known to those skilled in the art.

Depending on the expression system and host selected, the polypeptide of the present invention may be produced by growing host cells transformed by an expression vector described above under conditions whereby the polypeptide of interest is expressed. The polypeptide is then isolated from the host cells and purified. If the expression system secretes the polypeptide into growth media, the polypeptide can be purified directly from the media. If the polypeptide is not secreted, it is isolated from cell lysates or recovered from the cell membrane fraction. Where the polypeptide is localized to the cell surface, whole cells or isolated membranes can be used as an assayable source of the desired gene product. Polypeptide expressed in bacterial hosts such as *E. coli* may require isolation from inclusion bodies and refolding. Where the mature protein has a very hydrophobic region which leads to an insoluble product of overexpression, it may be desirable to express a truncated protein in which the hydrophobic region has been deleted. The selection of the appropriate growth conditions and recovery methods are within the skill of the art.

The polypeptide can be recovered and purified from recombinant cell cultures by methods including ammonium sulphate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. Polypeptides of the invention may also include an initial methionine amino acid residue.

Preparation of the RNase P RNA Component:

The RNase P RNA molecules are prepared by run-off *in vitro* transcription using T7 RNA polymerase as according to standard conditions - usually as recommended by the supplier, e.g., Promega. The plasmid is linearized with an appropriate restriction enzyme generating a linear dsDNA comprising the full length gene encoding the RNase P RNA. The RNA is purified either from a preparative denaturing acrylamide gel or is precipitated prior to use in *in vitro* cleavage assays. The substrates for the RNase P RNA and the RNA complexed with its protein (RNase P protein) can be obtained by *in vitro* transcription of cloned genes. Useful substrates included but are not limited to pre-tRNA^{Met} or *E. coli* or *B. subtilis* pre-4.5S molecules and may be expressed using an *in vitro* transcription system directed by T7 RNA polymerase as described above. The RNA can also be prepared by automated synthesis.

Antagonists and agonists - assays and molecules

This invention provides a method of screening drugs to identify those which interfere with the RNA portion, the protein portion and/ or the intact RNA/protein complex of the RNase P described herein, which method comprises measuring the interference of the activity of the protein and/or RNA by a test drug. For example since the RNA portion selected has a catalytic activity, after suitable purification and formulation the activity of the RNA can be followed by

its ability to convert its natural or synthetic RNA substrates. By incorporating different chemically synthesized test compounds or natural products into such an assay of enzymatic activity one is able to detect those additives which compete with the natural or synthetic substrate or otherwise inhibit enzymatic activity.

Polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics. See, e.g., Coligan *et al.*, *Current Protocols in Immunology* 1(2): Chapter 5 (1991).

The invention also provides a method of screening compounds to identify those which enhance (agonist) or block (antagonist) the action of RNaseP polypeptides or polynucleotides, particularly those compounds that are bacteriostatic

and/or bacteriocidal. The method of screening may involve high-throughput techniques. For example, to screen for agonists or antagonists, a synthetic reaction mix, a cellular compartment, such as a membrane, cell envelope or cell wall, or a preparation of any thereof, comprising RNaseP polypeptide and a labeled substrate or ligand of such polypeptide is incubated in the absence or the presence of a candidate molecule that may be a RNaseP agonist or antagonist. The ability of the candidate molecule to agonize or antagonize the RNaseP polypeptide is reflected in decreased binding of the labeled ligand or decreased production of product from such substrate. Molecules that bind gratuitously, *i.e.*, without inducing the effects of RNaseP polypeptide are most likely to be good antagonists. Molecules that bind well and increase the rate of product production from substrate are agonists. Detection of the rate or level of production of product from substrate may be enhanced by using a reporter system. Reporter systems that may be useful in this regard include but are not limited to colorimetric labeled substrate converted into product, a reporter gene that is responsive to changes in RNaseP polynucleotide or polypeptide activity, and binding assays known in the art.

Another example of an assay for RNaseP antagonists is a competitive assay that combines RNaseP and a potential antagonist with RNaseP-binding molecules, recombinant RNaseP binding molecules, natural substrates or ligands, or substrate or ligand mimetics, under appropriate conditions for a competitive inhibition assay. RNaseP can be labeled, such as by radioactivity or a colorimetric compound, such that the number of RNaseP molecules bound to a binding molecule or converted to product can be determined accurately to assess the effectiveness of the potential antagonist.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polynucleotide or polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the same sites on a binding molecule, such as a binding molecule, without inducing RNaseP-induced activities, thereby preventing the action of RNaseP by excluding RNaseP from binding.

Potential antagonists include a small molecule that binds to and occupies the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic molecules, peptides or peptide-like molecules. Other potential antagonists include antisense molecules (see Okano, *J. Neurochem.* 56: 560 (1991); *OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION*, CRC Press, Boca Raton, FL (1988), for a description of these molecules). Preferred potential antagonists include compounds related to and variants of RNaseP.

Each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein, upon expression, can be used as a target for the screening of antibacterial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to control the expression of the coding sequence of interest.

The invention also provides the use of the polypeptide, polynucleotide or inhibitor of the invention to interfere with the initial physical interaction between a pathogen and mammalian host responsible for sequelae of infection. In particular the molecules of the invention may be used: in the prevention of adhesion of bacteria, in particular gram positive bacteria, to mammalian extracellular matrix proteins on in-dwelling devices or to extracellular matrix proteins in wounds; to block RNaseP protein-mediated mammalian cell invasion by, for example, initiating phosphorylation of mammalian tyrosine kinases (Rosenshine *et al.*, *Infect. Immun.* 60:2211 (1992); to block bacterial adhesion between mammalian extracellular matrix proteins and bacterial RNaseP proteins that mediate tissue damage and; to block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques.

The antagonists and agonists of the invention may be employed, for instance, to inhibit and treat disease, such as, infections of the upper respiratory tract (e.g., otitis media, bacterial tracheitis, acute epiglottitis, thyroiditis), lower respiratory (e.g., empyema, lung abscess), cardiac (e.g., infective endocarditis), gastrointestinal (e.g., secretory diarrhoea, splenic abscess, retroperitoneal abscess), CNS (e.g., cerebral abscess), eye (e.g., blepharitis, conjunctivitis, keratitis, endophthalmitis, preseptal and orbital cellulitis, dacryocystitis), kidney and urinary tract (e.g., epididymitis, intrarenal and perinephric abscess, toxic shock syndrome), skin (e.g., impetigo, folliculitis, cutaneous abscesses, cellulitis, wound infection, bacterial myositis) bone and joint (e.g., septic arthritis, osteomyelitis).

HTP Screening Strategies:

Assays can be developed to detect compounds that inhibit RNase P directed cleavage of RNA substrates. Several possible assay formats are suitable for HTP screening based upon the ability to incorporate labels within the RNA in a site-specific fashion by chemical synthesis. The conventional radioactive-based format is preferred, while a homogeneous fluorescence-based format is useful for subsequent follow-up of lead compounds. The use of both formats is contemplated by this invention.

Functional RNase P Assay

Biotin is introduced in to an appropriate position within the RNA substrate and the 5' terminus labeled with ³²P. The substrate is linked via streptavidin within a 96-well plate. Following RNase P dependent hydrolysis of the substrate, the radiolabelled 5' leader cleavage product is released into the bulk solution phase, and subject to scintillation counting. Alternatively, the RNA substrate is bound to a streptavidin-coated flashplate such that the release of the radioactive 6-mer into the solution phase results in a decrease in signal. This has the advantage that it is a homogeneous, continuous assay format and requires no additional manipulations after starting the assay. Both formats are useful in the practice of this invention because they use the same RNA substrate.

RNA Fragment Library Rescue

An effective approach for identifying compounds that interact with RNA is contemplated. The concept is based on the over-expression of a drug binding site that is recreated on an RNA fragment, which will sequester the drug and permit the continued functioning of the intact ribozyme (Figure 6). This approach has recently been described in the context of a search for ligands that bind ribosomal RNA (Howard, B-A, *et al.*, Biochem. Cell Bio. 73(11/12): 1161-1166 (1995)). Following selection the random RNA fragments that apparently present a minimal target structure for drug recognition, are incorporated into a protocol for rational drug design. Accordingly, random fragment libraries based on MI RNA will be generated and used in HTP screening to identify compounds that disrupt RNA/protein interaction. (Figure 7)

Cyclic peptide phage libraries:

The incorporation of conformational constraints into flexible lead compounds is a powerful strategy to increase lead potency and is particularly useful in the field of peptidomimetic design. (Al-Obeidi, F. *et al.*, J. Med. Chem. 32: 2555-2561(1989); Barker, P.L., *et al.*, J. Med. Chem. 35: 2040-2048 (1992)). Cyclization has been shown to increase the propensity for beta-turn formation in peptides the potential of which has been demonstrated by the identification of high-affinity ligands for streptavidin (Lee, M. S., *et al.*, FEBS Lett 359: 113-118 (1995)). In this case, cyclic peptide libraries were constructed with flanking cysteine residues to allow efficient disulfide bond formation and cyclization during phage assembly. The streptavidin bound crystal structures of two disulfide bridged cyclic peptides showed both peptides to be in beta-turn conformations (Kahn, M. (Guest, Ed., 1993) Tetrahedron 49, Symp. 50, 3433-3677).

Beta-turns are key recognition elements in many biological interactions therefore effort has been focused on the design of small constrained beta-turn mimics (Kahn, M. (Guest Ed., 1993) Tetrahedron 49, Symp. 50, 3433-3677). This approach, when applied to RNase P, could identify cyclic peptides suitable for peptide mimic synthesis as inhibitor molecules. A cyclic octapeptide phage display library may be constructed and used to identify peptides that interact with defined RNA domains.

Secondary EvaluationSELEX: Systematic Evolution of Ligands by Exponential Enrichment:

This approach may be employed in an attempt to identify RNA recognition motifs for RNase P RNA (M1-RNA) protein binding for structural analysis as an aid to rational drug design and the secondary evaluation of compounds identified via the HTP screens. The technology is based on the repeated selection and amplification of RNA fragments that specifically bind to a protein with high affinities (Szostak, J.W., TIBS 17: 89-93 (1992)). Fragment libraries based on the *S. aureus* and *E. coli* RNase P RNAs may be constructed for the *in vitro* synthesis of RNA fragments and the subsequent selection of molecules that bind their respective proteins. Chemical and enzymatic structure probing technologies may be employed in combination with protein/RNA protection studies to map the interactive sites. SELEX based on the resulting RNA fragment(s) may be further exploited to determine the minimal structural requirements for

RNA recognition.

Disruption of RNase P Assembly

The identification of a protein/RNA-fragment pair permits the development of a screen for compounds that disrupt their assembly. Drug induced disruption of labeled RNA bound to immobilized protein (biotin/streptavidin) would result in the concomitant decrease/loss of the signal generated by the presence of the RNA (Figure 6).

RNA/Drug Interactions

RNase P RNA fragments that confer drug resistance (RNA Fragment Rescue Library *supra*) may be sequenced and expressed *in vitro* for chemical and enzymatic structure probing in the presence and absence of the drug in an attempt to map the binding site. SELEX may be applied to lead compounds in an attempt to identify the minimal structural requirements for drug binding.

RNase P Substrates

Minimal RNA substrates may be chemically synthesized for HTP screening including both pre-tRNA and pre-4.5S RNA derivatives (e.g. Figure 8). RNA-ligand interactions involving ribose 2'-hydroxyl groups of specific nucleotides may be probed via chemical synthesis of the appropriately modified RNA fragment. In order to retain the C^{3'}-*endo* configuration characteristic of ribonucleotides, 2'-methoxy and 2'-fluororibonucleotides analogues can be used, the latter being preferred on steric grounds. Nucleotides lacking a 2'-substituent adopt the undesired C^{2'}-*endo* configuration.

The invention also relates to inhibitors identified by any of the techniques described herein. Because of the enzymatic nature of RNase P action, it is appreciated that inhibitors may be identified which act as transition state mimics, inhibitors of product release or inhibitors of substrate binding.

Diagnostic Assays

This invention is also related to the use of the RNaseP polynucleotides of the invention for use as diagnostic reagents. Detection of RNaseP in a eukaryote, particularly a mammal, and especially a human, will provide a diagnostic method for diagnosis of a disease. Eukaryotes (herein also "individual(s)"), particularly mammals, and especially humans, infected with an organism comprising the RNaseP gene may be detected at the nucleic acid level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from an infected individual's cells and tissues, such as bone, blood, muscle, cartilage, and skin. Genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification technique prior to analysis. RNA or cDNA may also be used in the same ways. Using amplification, characterization of the species and strain of prokaryote present in an individual, may be made by an analysis of the genotype of the prokaryote gene. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the genotype of a reference sequence. Point mutations can be identified by hybridizing amplified DNA to labeled RNaseP polynucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in the electrophoretic mobility of the DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing. See, e.g., Myers et al., *Science*, 230: 1242 (1985). Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase and S1 protection or a chemical cleavage method. See, e.g., Cotton et al., *Proc. Natl. Acad. Sci., USA*, 85: 4397-4401 (1988).

Cells carrying mutations or polymorphisms in the gene of the invention may also be detected at the DNA level by a variety of techniques, to allow for serotyping, for example. For example, RT-PCR can be used to detect mutations. It is particularly preferred to use RT-PCR in conjunction with automated detection systems, such as, for example, GeneScan. RNA or cDNA may also be used for the same purpose, PCR or RT-PCR. As an example, PCR primers complementary to a nucleic acid encoding RNaseP can be used to identify and analyze mutations. Examples of representative primers are shown in the Examples. The invention further provides these primers with 1, 2, 3 or 4 nucleotides removed from the 5' and/or the 3' end. These primers may be used for, among other things, amplifying RNaseP DNA isolated from a sample derived from an individual. The primers may be used to amplify the gene isolated from an infected individual such that the gene may then be subject to various techniques for elucidation of the DNA sequence. In this way, mutations in the DNA sequence may be detected and used to diagnose infection and to serotype and/or classify the infectious agent.

The invention further provides a process for diagnosing, disease, preferably bacterial infections, more preferably

infections by *Staphylococcus aureus*, and most preferably disease, such as, infections of the upper respiratory tract (e.g., otitis media, bacterial tracheitis, acute epiglottitis, thyroiditis), lower respiratory (e.g., empyema, lung abscess), cardiac (e.g., infective endocarditis), gastrointestinal (e.g., secretory diarrhoea, splenic abscess, retroperitoneal abscess), CNS (e.g., cerebral abscess), eye (e.g., blepharitis, conjunctivitis, keratitis, endophthalmitis, preseptal and orbital cellulitis, dacryocystitis), kidney and urinary tract (e.g., epididymitis, intrarenal and perinephric abscess, toxic shock syndrome), skin (e.g., impetigo, folliculitis, cutaneous abscesses, cellulitis, wound infection, bacterial myositis) bone and joint (e.g., septic arthritis, osteomyelitis), comprising determining from a sample derived from an individual a increased level of expression of polynucleotide having the sequence of Table 1 [SEQ ID NO: 1]. Increased or decreased expression of RNaseP polynucleotide can be measured using any one of the methods well known in the art for the quantitation of polynucleotides, such as, for example, amplification, PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

In addition, a diagnostic assay in accordance with the invention for detecting over-expression of RNaseP protein compared to normal control tissue samples may be used to detect the presence of an infection, for example. Assay techniques that can be used to determine levels of a RNaseP protein, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Antibodies

The polypeptides of the invention or variants thereof, or cells expressing them can be used as an immunogen to produce antibodies immunospecific for such polypeptides. "Antibodies" as used herein includes monoclonal and polyclonal antibodies, chimeric, single chain, simianized antibodies and humanized antibodies, as well as Fab fragments, including the products of an Fab immunoglobulin expression library.

The Fab fragment may also be prepared from its parent monoclonal antibody by enzyme treatment, for example using papain to cleave the Fab portion from the Fc portion.

Antibodies generated against the polypeptides of the invention can be obtained by administering the polypeptides or epitope-bearing fragments, analogues or cells to an animal, preferably a nonhuman, using routine protocols. The antibody so obtained will then bind the polypeptides itself. In this manner, even a sequence encoding only a fragment of the polypeptides can be used to generate antibodies binding the whole native polypeptides. Such antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide. For preparation of monoclonal antibodies, any technique known in the art that provides antibodies produced by continuous cell line cultures can be used. Examples include various techniques, such as those in Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, pg. 77-96 in *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc. (1985), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, *et al.*, 1985, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96).

The hybridomas are screened to select a cell line with high binding affinity and favorable cross reaction with other staphylococcal species using one or more of the original polypeptide and/or the fusion protein. The selected cell line is cultured to obtain the desired Mab.

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies.

Alternatively phage display technology may be utilized to select antibody genes with binding activities towards the polypeptide either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing anti-RNaseP or from naive libraries (McCafferty, J. *et al.*, (1990), *Nature* 348, 552-554; Marks, J. *et al.*, (1992) *Bio-technology* 10, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. *et al.*, (1991) *Nature* 352, 624-628).

The antibody should be screened again for high affinity to the polypeptide and/or fusion protein.

As mentioned above, a fragment of the final antibody may be prepared.

The antibody may be either intact antibody of M_r approx 150,000 or a derivative of it, for example a Fab fragment or a Fv fragment as described in Skerra, A and Pluckhun, A., *Science* 240:1038-1040 (1988). If two antigen binding domains are present each domain may be directed against a different epitope - termed 'bispecific' antibodies.

If two antigen binding domains are present each domain may be directed against a different epitope - termed 'bispecific' antibodies.

In particular derivatives which are slightly longer or slightly shorter than the native protein or polypeptide fragment of the present invention may be used. In addition, polypeptides in which one or more of the amino acid residues are modified may be used. Such peptides may, for example, be prepared by substitution, addition, or rearrangement of amino acids or by chemical modification thereof. All such substitutions and modifications are generally well known to those skilled in the art of peptide chemistry.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptides to purify the polypeptides by affinity chromatography.

Thus, among others, antibodies against RNaseP- polypeptide may be employed to treat infections, particularly bacterial infections and especially disease, such as, infections of the upper respiratory tract (e.g., otitis media, bacterial tracheitis, acute epiglottitis, thyroiditis), lower respiratory (e.g., empyema, lung abscess), cardiac (e.g., infective endocarditis), gastrointestinal (e.g., secretory diarrhoea, splenic abscess, retroperitoneal abscess), CNS (e.g., cerebral abscess), eye (e.g., blepharitis, conjunctivitis, keratitis, endophthalmitis, preseptal and orbital cellulitis, dacryocystitis), kidney and urinary tract (e.g., epididymitis, intrarenal and perinephric abscess, toxic shock syndrome), skin (e.g., impetigo, folliculitis, cutaneous abscesses, cellulitis, wound infection, bacterial myositis) bone and joint (e.g., septic arthritis, osteomyelitis).

Preferably the antibody is prepared by expression of a DNA polymer encoding said antibody in an appropriate expression system such as described above for the expression of polypeptides of the invention. The choice of vector for the expression system will be determined in part by the host, which may be a prokaryotic cell, such as *E. coli* (preferably strain B) or *Streptomyces sp.* or a eukaryotic cell, such as a mouse C127, mouse myeloma, human HeLa, Chinese hamster ovary, filamentous or unicellular fungi or insect cell. The host may also be a transgenic animal or a transgenic plant (for example, as described in Hiatt, A. *et al.*, *Nature* 340:76-78(1989). Suitable vectors include plasmids, bacteriophages, cosmids and recombinant viruses, derived from, for example, baculoviruses and vaccinia.

Polypeptide variants include antigenically, epitopically or immunologically equivalent variants that form a particular aspect of this invention. The term "antigenically equivalent derivative" as used herein encompasses a polypeptide or its equivalent which will be specifically recognized by certain antibodies which, when raised to the protein or polypeptide according to the invention, interfere with the immediate physical interaction between pathogen and mammalian host. The term "immunologically equivalent derivative" as used herein encompasses a peptide or its equivalent which when used in a suitable formulation to raise antibodies in a vertebrate, the antibodies act to interfere with the immediate physical interaction between pathogen and mammalian host.

The polypeptide, such as an antigenically or immunologically equivalent derivative or a fusion protein thereof is used as an antigen to immunize a mouse or other animal such as a rat or chicken. The fusion protein may provide stability to the polypeptide. The antigen may be associated, for example by conjugation, with an immunogenic carrier protein for example bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH). Alternatively a multiple antigenic peptide comprising multiple copies of the protein or polypeptide, or an antigenically or immunologically equivalent polypeptide thereof may be sufficiently antigenic to improve immunogenicity so as to obviate the use of a carrier.

Preferably, the antibody or variant thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be "humanized"; where the complementarity determining region(s) of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody, for example as described in Jones, P. *et al.* (1986), *Nature* 321, 522-525 or Tempest *et al.*, (1991) *Biotechnology* 9, 266-273. The humanized monoclonal antibody, or its fragment having binding activity, form a particular aspect of this invention.

The modification need not be restricted to one of "humanization"; other primate sequences (for example Newman, R. *et al.*, *Biotechnology* 10:1455-1460 (1992)) may also be used.

The use of a polynucleotide of the invention in genetic immunization will preferably employ a suitable delivery method such as direct injection of plasmid DNA into muscles (Wolff *et al.*, *Hum Mol Genet* 1992, 1:363, Manthorpe *et al.*, *Hum. Gene Ther.* 1993:4, 419), delivery of DNA complexed with specific protein carriers (Wu *et al.*, *J Biol Chem.* 1989: 264, 16985), coprecipitation of DNA with calcium phosphate (Benvenisty & Reshef, *PNAS USA*, 1986:83,9551), encapsulation of DNA in various forms of liposomes (Kaneda *et al.*, *Science* 1989:243,375), particle bombardment (Tang *et al.*, *Nature* 1992, 356:152, Eisenbraun *et al.*, *DNA Cell Biol* 1993, 12:791) and *in vivo* infection using cloned retroviral vectors (Seeger *et al.*, *PNAS USA* 1984:81,5849).

Vaccines

Another aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal which comprises inoculating the individual with RNaseP, or a fragment or variant thereof, adequate to produce antibody and/ or T cell immune response to protect said individual from infection, particularly bacterial infection and most particularly *Staphylococcus aureus* infection. Also provided are methods whereby such immunological response slows bacterial replication. Yet another aspect of the invention relates to a method of inducing immunological response in an individual which comprises delivering to such individual a nucleic acid vector to direct expression of RNaseP, or a fragment or a variant thereof, for expressing RNaseP, or a fragment or a variant thereof *in vivo* in order to induce an immunological response, such as, to produce antibody and/ or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said individual from disease, whether that disease is already established within the individual or not. One way of administering the gene is by accelerating it into the desired cells as a coating on particles or otherwise. Such nucleic acid vector may comprise DNA, RNA, a modified

nucleic acid, or a DNA/RNA hybrid.

A further aspect of the invention relates to an immunological composition which, when introduced into an individual capable or having induced within it an immunological response, induces an immunological response in such individual to a RNaseP or protein coded therefrom, wherein the composition comprises a recombinant RNaseP or protein coded therefrom comprising DNA which codes for and expresses an antigen of said RNaseP or protein coded therefrom. The immunological response may be used therapeutically or prophylactically and may take the form of antibody immunity or cellular immunity such as that arising from CTL or CD4+ T cells.

A RNaseP polypeptide or a fragment thereof may be fused with co-protein which may not by itself produce antibodies, but is capable of stabilizing the first protein and producing a fused protein which will have immunogenic and protective properties. Thus fused recombinant protein, preferably further comprises an antigenic co-protein, such as lipoprotein D from *Hemophilus influenzae*, Glutathione-S-transferase (GST) or beta-galactosidase, relatively large co-proteins which solubilize the protein and facilitate production and purification thereof. Moreover, the co-protein may act as an adjuvant in the sense of providing a generalized stimulation of the immune system. The co-protein may be attached to either the amino or carboxy terminus of the first protein.

Provided by this invention are compositions, particularly vaccine compositions, and methods comprising the polypeptides or polynucleotides of the invention and immunostimulatory DNA sequences, such as those described in Sato, Y. *et al.* Science 273: 352 (1996).

Also, provided by this invention are methods using the described polynucleotide or particular fragments thereof which have been shown to encode non-variable regions of bacterial cell surface proteins in DNA constructs used in such genetic immunization experiments in animal models of infection with *Staphylococcus aureus* will be particularly useful for identifying protein epitopes able to provoke a prophylactic or therapeutic immune response. It is believed that this approach will allow for the subsequent preparation of monoclonal antibodies of particular value from the requisite organ of the animal successfully resisting or clearing infection for the development of prophylactic agents or therapeutic treatments of bacterial infection, particularly *Staphylococcus aureus* infection, in mammals, particularly humans.

The polypeptide may be used as an antigen for vaccination of a host to produce specific antibodies which protect against invasion of bacteria, for example by blocking adherence of bacteria to damaged tissue. Examples of tissue damage include wounds in skin or connective tissue caused, e.g., by mechanical, chemical or thermal damage or by implantation of indwelling devices, or wounds in the mucous membranes, such as the mouth, mammary glands, urethra or vagina.

The invention also includes a vaccine formulation which comprises an immunogenic recombinant protein of the invention together with a suitable carrier. Since the protein may be broken down in the stomach, it is preferably administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, or intradermal. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

While the invention has been described with reference to certain RNaseP protein, it is to be understood that this covers fragments of the naturally occurring protein and similar proteins with additions, deletions or substitutions which do not substantially affect the immunogenic properties of the recombinant protein.

Compositions, kits and administration

The invention also relates to compositions comprising the polynucleotide or the polypeptides discussed above or their agonists or antagonists. The polypeptides of the invention may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to a subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of a polypeptide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should suit the mode of administration. The invention further relates to diagnostic and pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

Polypeptides and other compounds of the invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes among others.

In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

Alternatively the composition may be formulated for topical application for example in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for example cream or ointment bases, and ethanol or oleyl alcohol for lotions. Such carriers may constitute from about 1% to about 98% by weight of the formulation; more usually they will constitute up to about 80% by weight of the formulation.

For administration to mammals, and particularly humans, it is expected that the daily dosage level of the active agent will be from 0.01 mg/kg to 10 mg/kg, typically around 1 mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual and will vary with the age, weight and response of the particular individual. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

In-dwelling devices include surgical implants, prosthetic devices and catheters, i.e., devices that are introduced to the body of an individual and remain in position for an extended time. Such devices include, for example, artificial joints, heart valves, pacemakers, vascular grafts, vascular catheters, cerebrospinal fluid shunts, urinary catheters, continuous ambulatory peritoneal dialysis (CAPD) catheters.

The composition of the invention may be administered by injection to achieve a systemic effect against relevant bacteria shortly before insertion of an in-dwelling device. Treatment may be continued after surgery during the in-body time of the device. In addition, the composition could also be used to broaden perioperative cover for any surgical technique to prevent bacterial wound infections, especially *Staphylococcus aureus* wound infections.

Many orthopaedic surgeons consider that humans with prosthetic joints should be considered for antibiotic prophylaxis before dental treatment that could produce a bacteremia. Late deep infection is a serious complication sometimes leading to loss of the prosthetic joint and is accompanied by significant morbidity and mortality. It may therefore be possible to extend the use of the active agent as a replacement for prophylactic antibiotics in this situation.

In addition to the therapy described above, the compositions of this invention may be used generally as a wound treatment agent to prevent adhesion of bacteria to matrix proteins exposed in wound tissue and for prophylactic use in dental treatment as an alternative to, or in conjunction with, antibiotic prophylaxis.

Alternatively, the composition of the invention may be used to bathe an indwelling device immediately before insertion. The active agent will preferably be present at a concentration of 1 µg/ml to 10 mg/ml for bathing of wounds or indwelling devices.

A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response. A suitable unit dose for vaccination is 0.5-5 microgram/kg of antigen, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks. With the indicated dose range, no adverse toxicological effects will be observed with the compounds of the invention which would preclude their administration to suitable individuals.

Each reference disclosed herein is incorporated by reference herein in its entirety. Any patent application to which this application claims priority is also incorporated by reference herein in its entirety.

Examples

The present invention is further described by the following examples. The examples are provided solely to illustrate the invention by reference to specific embodiments. These exemplifications, while illustrating certain specific aspects of the invention, do not portray the limitations or circumscribe the scope of the disclosed invention.

All examples were carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. Routine molecular biology techniques of the following examples can be carried out as described in standard laboratory manuals, such as Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

All parts or amounts set out in the following examples are by weight, unless otherwise specified.

Unless otherwise stated size separation of fragments in the examples below was carried out using standard techniques of agarose and polyacrylamide gel electrophoresis ("PAGE") in Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989) and numerous other references such as, for instance, by Goeddel et al., *Nucleic Acids Res.* 8: 4057 (1980).

Exempl 1. PCR with random primer

This method describes a rapid way to obtain additional sequence data from partial gene fragments independent of probing a genomic library.

Random sequencing of a *S. aureus* genomic library followed by sequence homology searching with the *B. subtilis* P-protein sequence resulted in the identification of a 324 base pair (herein "bp") fragment. The first 193 nucleotides of this fragment showed significant homology to the C-terminal half of the *B. subtilis* P-protein and other prokaryotic RNase P proteins. The N-terminal domain of the putative *S. aureus* RNase P protein was putatively missing.

Deduced *S. aureus* RNase P protein sequence versus *B. subtilis* shows 58.7% similarity, 34.9% identity (see Figure 2 for an alignment of RNase P protein sequences).

A novel two step PCR with two different reverse primers complementary to the known sequence in position 15-39 and 194-215 respectively (shown below, Table 2) as well as random hexamer primers (Gibco) was used to obtain the complete 5' sequence of the *S. aureus spp* gene. *S. aureus* genomic DNA partially digested with Hind III or Pst I served as a template. Primer #1 (position 194-215) annealed just downstream of the stop codon TAA and primer #2 (position 15-39) close to the end of the known sequence (see Table 2).

Table 2**Primer Position**

```

5' GAATT CCGTG TTAAG AAACA AGATT AAAAG AGCAA TACGT GAAAA TTICA AAGTA CATAA GTCGN
   3' C TTTGT TCTAA TTTTC TCGTT ATGC 5' (SEQ ID NO: 7)
ATATA TTGGN CAAAG ATATT ATTGN TATAG NAAGA CAGGC AGCTA AAGAT ATGAC GACTT TACAA

ATACA GNATA GTCTT GAGCA COTNC TTAAG ATTGG CAAAG TTTT AATAA AAAGA TTAAG TAAGG
   3' CC
ATAGG GTAGG GAAAG AAAAC ATTAA CCCCT CAACG CATCC CGAAG TCTTA CCTCA GACAA ACGTT
TATCC CATCC CTTAC TTTTG 5' (SEQ ID NO: 8)
AGACT GACCC TAGGG TTAAG ACTTG GCCCN AGGGN TATNN TAACT TACTT TAAAA TGTTT TCAC 3'

```

In the first step, the reverse strand encoding the C-terminus and upstream the unknown N-terminal sequence of the protein was amplified using primer #1, resulting in single stranded products as diagrammed in Figure 9. This amplification of the desired sequence favoured the binding of random primers to that sequence in the next step.

In the second step, 32P-end-labeled primer #2 and random hexamers were added and the annealing temperature was dropped to 25°C to allow the short random primers to anneal to the DNA.

The PCR is expected to generate many different fragments, however the major products should be those primed by primer #2 and a random primer (see below). Only these products were of interest as they should encode the N-terminal half of the protein and they could be monitored since they would be radiolabeled. After separation of the PCR products by gel electrophoresis and exposure to film, a limited number of radiolabeled bands should be visible on the autoradiogram. This method is diagrammed in Figure 10. These fragments can subsequently be cloned into a suitable vector, e.g., pUC19, and sequenced using routine methods.

Example 2. PCR of the *spp* 5' domain with random primers

The 5' half of the *spp* gene was amplified in a two step PCR reaction as described herein. The PCR products were cleaned using a QIAquick PCR purification column and recovered in 50 µl water. The sample was concentrated to a final volume of 25 µl under vacuum in a SpeedVac® SC110 (Savant). 20µl were loaded onto a 1.6 mm, 8% polyacrylamide gel containing 7M urea and analyzed by gel electrophoresis. The gel was stained with ethidium bromide (1 µg/ml in H₂O) to visualize the DNA and photographed. The gel was then dried and exposed to film.

There were several defined bands visible both on the ethidium bromide stained gel as well as on the autoradiogram. They ranged in size between 50 bp to over 2000 bp, some bands being more intense than others. The most prominent bands were approximately 50bp, 150bp, 400bp and 800bp in size.

Example 3. Shotgun cloning of the PCR fragments: Cloning and sequencing of the PCR fragments

The PCR fragments were blunt end cloned into SmaI cut pUC19. Ligations of 100 ng vector DNA with 1.0 µl and 1.5 µl of PCR products were performed at 16°C overnight. *E. coli* XL1 blue cells (a-complementing) were transformed by electroporation and plated onto selective plates containing IPTG and x-gal (for blue/white screening). Transformation of these cells resulted in approximately 8.5% white colonies in the presence of PCR products, whereas

only 4.6% white colonies could be found in the control ligation without insert. White colonies had lost the ability to form the β -galactosidase holoenzyme as a result of a reading frame disruption in the *lacZ'* gene and were no longer able to cleave x-gal. The reading frame was either disrupted by an insert (in the presence of PCR products) or exonuclease activity of the restriction enzyme preparation, destroying the Sma I site of the vector by cleaving terminal nucleotides and resulting in frame shifts after ligation.

18 white colonies were picked and grown up as 2 ml cultures to prepare plasmids. A sample of each plasmid was linearized with AccI restriction endonuclease (cuts in the multi cloning site) and run on a 1% TAE agarose gel alongside the original pUC19.

The results showed that 7 plasmids released a short fragment upon restriction digest. As there is only one AccI site in pUC19, these recombinant plasmids must contain an insert with a restriction site for this enzyme. 3 plasmids appeared larger than pUC 19 but did not release a fragment. The other 6 plasmids were of the same size as pUC19 and did not appear to have an insert.

The 10 plasmids containing an insert were sequenced both with the forward and reverse pUC/M13 sequencing primers.

5 of the 6 plasmids that released a fragment after AccI digest harboured an 213bp insert that started with the primer #2 sequence and the known 14bp followed by a sequence of 174 nucleotides that was common to all of the plasmids (Sequence 1, Table 3). In one of the 6 plasmids the insert (Sequence 2, Table 3) was slightly different: the first 137 nucleotides were in common with Sequence 1 (dotted line), however the insert sequence extended beyond this and deviated from Sequence 1 (see Table 3):

Table 3

Aplification Product Sequence 1, 2 and 3

Sequence 1 [SEQ ID NO: 9]

5' GTTTT CATT C C T A C C T A T C C T T A C T T A T T A T T A A A A A
 G A A T G C A G A T T T T C A G A G A A T A T A T A A A A A A G G T C A T T C T G T A G C
 C A A C A G A C A A T T T G T T G T A T A C A C T T G T A A T A A T A A G A A A T A G A
 C C A T T T T C G C T T A G G T A T T A G T G T T T C T A A A A A C T A G G T A A T G C
 A G T G T T A A G A A C A A G A T T A A A A G A G C A A T A C G T 3'

Sequence 2 [SEQ ID NO: 10]

5' ACGAT AAATA ACAGT ATGTT ATTGG AAAAA GTCTA CCGAA TTAAA
 AAGAA TGCAG ATTTT CAGAG AATAT ATAAA AAAGG TCATT CTGTA GCCAA
 CAGAC AATTT GTTGT ATACA CTTGT AATAA TAAAG AAATA GACCA TTTTC
 GCTTA GGTAT TAGTG TTTCT AAAAA ACTAG GTAAT GCAGT GTTAA GAACA
 AGATT AAAAG AGCAA TACGT 3'

Sequence 3 [SEQ ID NO: 11]

5' . . . GAAAA TTTC AAGTA CATAA GTCGC ATATA TTGGC CAAAG ATTGT
 AATAG CAAGA CAGCC AGCTA AAGAT ATGAC GACTT TACAA ATACA GAATA
 GTCTT GAGCA CGTAC TTAAG ATTGC CAAAG TTTT AATAA AAAGA TTAAG
TAAAG ATAGG GTAGG GAATG AAAAC 3'
 stop codon (underlined)

Primer #2 sequence is shown in bold and also the known 14 bases underlined (3 changes to known sequence in italic).

The predicted translation products of Sequence 1 and 2 showed significant homology to *B. subtilis* P-protein and indicated that these inserts contain the full sequence information for the 5' domain of the *spp* gene.

In 2 plasmids the insert started with the sequence of primer #1 followed by 3' half of the *spp* gene (Sequence 3, Table 3). This sequence was already known but the new sequence showed 9 base changes (shown in bold) in comparison to the original sequence data obtained by random sequencing of the *S. aureus* library.

2 plasmids contained inserts with different sequences that are a result of unspecific priming during PCR.

Example 4 Analysis of the sequence data

The newly obtained sequence data was drawn together and predicted to represent the complete sequence of the *spp* structural gene. The reading frame was determined from a partial sequence of the 3' domain. According to this reading frame the start codon ATG could be identified 174bp upstream of primer #2 sequence (see above). The total length of the gene is 354bp.

The nucleotide sequence of the *spp* gene is shown in Table 1 [SEQ ID NO: 2]; the start codon (ATG) and stop codon (TAA) are both shown in bold. The deduced translation product of the *spp* gene (SP protein) is 117 amino acids is also shown in Table 1 [SEQ ID NO: 2].

The translated sequence was aligned with known amino acid sequences of the RNase P protein from 5 different prokaryotes. These alignments are shown in Figure 2. The *S. aureus* SP protein sequence matched the *B. subtilis* P-protein sequence well, significant homology to the *B. subtilis* protein.

Example 5. PCR of the *spp* gene

The newly obtained sequence data was used to design PCR primers for the amplification of the full length *spp* structural gene.

The PCR resulted in a single band of approximately 360bp with a total DNA yield of 2 µg as determined by spectrophotometrical measurement. The DNA fragment was cleaned using a QIAquick PCR purification column and recovered in 80 µl water.

Example 6. Cloning *spp* into pMalc2

The pMalc2 vector provides a method for expressing and purifying a protein produced from a cloned gene (New England BioLabs) (see also. Guan, C., Li, P., Riggs, P.D. and Inouye, H. (1987) Gene 67, 21-30; Maina, C.V. et al.

(1988) Gene 74, 365-373; Riggs, P. in Ausubel, F.M. et al. (eds) Current prot. in Molecular Biol. (1992); Kellerman and Ferenci (1982) Methods in Enzymol. 90, 459-463; Yanisch-Perron, C. et al (1985) Gene 33, 103-119; Zagursky, R.J. and Berman, M.L. (1984) Gene 27, 183-191, regarding the vector and its uses).

The cloned gene is inserted downstream from the *malE* gene of *E. coli*, which encodes the maltose-binding-protein (MBP), resulting in the expression of an MBP fusion protein. The method uses the strong "tac" promoter under the control of the lac repressor and a one step purification of the fusion protein using MBP's affinity for maltose. Unique restriction sites are available between *malE* and *lacZa* for inserting the gene of interest, allowing blue/white screening of transformants. The vector also contains a recognition site for the specific Factor Xa protease located just 5' to the polylinker. This allows MBP to be cleaved after purification without leaving any vector derived residues attached to the protein of interest when the gene is cloned into the *XmnI* site.

Example 7. Ligation of the *spp* PCR product into pMalc2

1 µg of the protein fusion vector pMalc2 was cut with *XmnI* and *BamHI* restriction endonucleases to enable directional cloning of the *spp* structural gene. To remove the small 15 bp insert the digest was cleaned using QIAquick PCR purification column and the DNA recovered in 90 µl water. A sample was analysed by agarose gel electrophoresis to ensure complete digestion of the vector. After 30 min of incubation at 37°C with both enzymes the vector was completely cut.

The *spp* PCR product was digested with *BamHI* and cleaned using a QIAquick PCR purification column.

Vector and insert were ligated overnight in a 20 µl reaction at 16°C. The molar ratio of vector to insert was 1:5. To prepare the DNA for electroporation, the ligation was desalted using QIAquick Nucleotide Removal columns.

Example 8. Transformation of *E. coli* XL-1 Blue cells

40 µl electrocompetent *E. coli* XL-1 Blue cells were transformed with 2 µl desalted ligation reaction. Cells were recovered in 1ml SOC medium after electroporation. 50 µl of a 1:100 dilution were plated onto an LB plate containing ampicillin, IPTG and x-gal and incubated at 37°C overnight. 11 white and 1 blue colony were recovered. The 11 white colonies were grown up and the plasmids were isolated from the clones. Sequencing of the plasmids showed that 6 of the 11 plasmids contained the correct *spp* insert. The resulting plasmid was denoted pMalc2::spp, and the fusion protein product MBP-SPP.

Example 9 Growth of cells and fusion protein induction in *E. coli* XL-1 Blue pMalc2::spp

A 1 liter culture of *E. coli* XL-1 Blue pMalc2::spp was grown at 37°C, 220 rpm and expression of the MBP-SPP fusion induced by adding 1mM IPTG at an A^{600} of 0.6. Induction was continued for 2 hours. The A^{600} was monitored over that time period to ensure that cells were still growing during induction. 1ml samples were removed from the culture after 0.5, 1.0 1.5 and 2 hours, the cells pelleted and the pellet resuspended in 100 µl SDS gel loading buffer. 3 µl of the samples were analysed by SDS-PAGE with Coomassie staining.

The cells were not effected by the overexpression of the fusion protein, they kept growing and after 2 hours of IPTG induction reached an A^{600} of 1.43.

Protein extracts from samples were separated by SDS-PAGE and transferred onto a nitro-cellulose membrane by Western transfer. The MBP-SPP fusion protein could be immuno-detected by incubation of the membrane with anti-MBP rabbit serum and HRP-coupled anti-rabbit antibodies followed by ECL detection. The Western Blot revealed that 2 proteins were induced upon IPTG addition. The major protein was of the expected size of approximately 56kD, the second protein, produced in lower amounts, was 10-15 kD larger. Analysis of the *spp* sequence revealed an accumulation of AUA (ile) and AGA (arg) codons that are rare in *E. coli*. The normal frequency of AGA in *E. coli* is 0.27%, of AUA 0.51%. The *spp* sequence consists to 4.3% of AGA and 5.1% of AUA. These codons may cause frame shifts and other altered translation events when overexpressing a protein with a high abundance of rare codons.

Another *E. coli* strain was available that constitutively expresses the ArgU tRNA ($tRNA^{UCU}$) and IleX tRNA ($tRNA^{UAU}$) from the plasmid pRI952.

Example 10. Transformation of *E. coli* W3110 pRI952

50 µl electrocompetent *E. coli* W3110 pRI952 cells were transformed with 2 µl desalted ligation reaction. Cells were recovered in 1ml SOC medium after electroporation. 100 µl of the culture were plated onto selective LB plates (ampicillin, cam) containing IPTG and x-gal and incubated at 37°C overnight. Too many colonies grew on the plate and could not be counted. 5 white colonies were selected and their plasmids isolated. Sequencing of the plasmids showed that all contained the correct *spp* insert.

Example 11. Growth of cells and fusion prot in induction in *E. coli* W3110 pRI952 pMalc2::spp

The cells were grown and the expression of the fusion protein was induced as described above. The cells kept growing during the induction period from an A_{600} of 0.6 to 1.45. Samples were taken from the culture at 0.0, 0.5, 1.5 and 2.5 hours after IPTG addition and protein extracts analysed by SDS-PAGE with Coomassie staining, as well as Western transfer with immunodetection. Overexpression of the rare tRNAs IleX and ArgU resulted in only one MBP-SPP fusion protein being overexpressed during IPTG induction as detected by the anti-MBP serum.

Example 12. Purification of the MBP-SPP fusion by affinity chromatography

Following optimal induction, the cells (*E. coli* W3110 pRI952 pMalc2::spp) were harvested and disrupted by sonication after a freeze-thaw cycle to weaken the cell wall using known methods. The cell debris was removed by centrifugation and the protein concentration of the crude extract determined using the BIO-RAD Protein Assay. The total amount of protein in the crude extract was 86 mg. A sample of the crude extract was analysed by SDS-PAGE with Coomassie staining which showed that the fusion protein was released from the cells during sonication.

The crude extract was then pumped through the amylose resin column, to which the MBP-SPP fusion bound, and unbound protein was removed from the column by extensive washing with column buffer. The fusion protein was then eluted with column buffer supplemented with 10mM maltose using methods described herein. The eluate was collected as 3 ml fractions. Samples were removed during the purification procedure and analyzed by SDS-PAGE with Coomassie staining. Almost all of the fusion protein bound to the amylose resin, only trace amounts were found in the flow-through and no fusion protein was washed off the column. Fractions 4-11 contained reasonable amounts of protein and were pooled, yielding in 25 ml protein solution with a concentration of 1 mg/ml.

Example 13. Factor Xa cleavage of the fusion protein and separation of MBP and SPP

500 µg Factor Xa protease were added to the solution and the fusion protein was cleaved overnight at 4°C. A sample was analysed by SDS-PAGE before and after Factor Xa incubation to monitor complete cleavage. After 15 hours almost all the fusion protein was cleaved resulting in the 42 kD MBP and SPP of approximately 14kD.

The SP protein readily precipitated upon Factor Xa cleavage and could be separated from MBP by centrifugation. The SPP pellet was washed 3 times and then resuspended in 5ml column buffer. The MBP remained in solution and was found in the supernatant whereas the pellet consisted mainly of the SP protein and one minor contaminating protein, resulting in a reasonably pure preparation of SPP. The total SPP yield was 2.18 mg/liter culture. The SPP preparation was diluted with column buffer to a final concentration of 0.32 mg/ml. 1mM DTT was added to prevent intermolecular disulfide bond formation between cystein residues. The addition of 7M urea denatured and apparently completely resolubilized the protein. The protein was aliquoted, snap-frozen and stored at -80°C.

Example 14. *In vitro* transcription of the SP RNA

SP RNA was transcribed *in vitro* from the plasmid pSPR using T7 RNA polymerase. pSPR is a pUC19 derivative that harbours the *S. aureus* spr gene behind a T7 promotor. The plasmid was linearized with BamHI restriction endonuclease to enable run off transcription. The Ambion MEGAscript™ T7 kit was used for large scale transcription of the SP RNA. The transcription reaction was performed at 30°C to allow slow folding of the RNA into the correct conformation during synthesis. The RNA was cleaned under non-denaturing conditions using CLONTECH Chroma-Spin 30 columns and the RNA recovered in 40 µl water. The quality of the transcript was monitored by gel electrophoresis on a 5% polyacrylamide TBE gel with 7M urea. The RNA was visualized by UV-shadowing over a TLC plate.

In vitro transcription of SP RNA from pSPR resulted in a single RNA product of the expected size of 401 nucleotides. The total RNA yield was 140 µg (determined spectrophotometrically). After cleaning, the RNA was recovered in 40 µl water resulting in a 29 µM solution, that was stored at -20°C.

Example 15. *In vitro* transcription of *E. coli* ptRNA^{met} incorporating [α -³²P] UTP

E. coli ptRNA^{met}, a substrate for RNase P, was transcribed *in vitro* by T7 RNA polymerase (Epicenter) in the presence of [α -³²P] UTP to internally label the RNA. The DNA template pGem3Z-ptRNA was linearized with BstNI restriction endonuclease to allow run-off transcription. The transcription was performed in a 20 µl reaction at 37°C for 60min. To remove unincorporated nucleotides the 93nt RNA (sequence shown below) was cleaned under non-denaturing conditions using a CLONTECH Chroma-Spin 10 column. To calculate the specific activity of the RNA probe, 1ul samples were removed before and after cleaning, mixed with 5 ml scintillation cocktail and counted in a scintillation counter.

ptRNA^{met} in pGem-3Z, sequence of the transcript (93mer) 5'→3' [SEQ ID NO: 12]:

GGGCG AAUUC GCCUC GGCUA CGUAG CUCAG UUGGU UAGAG CACAU CACUC
 AUAUU GAUGG GGUCA CAGGU UCGAA UCCCG UCGUA GCCAC CAG

Transcription of ptRNA^{met} resulted in a 2μM (6 ng/μl) solution. 64.3% of the nucleotides were incorporated in the transcript and the specific activity was 2.171 10⁸ cpm/μg.

Example 16. 5' end-labeling of p6AT-1

The chemically synthesized 42nt RNA molecule p6AT-1 was 5' end-labeled with high specific activity [γ -³²P] ATP (6000μCi/μl) using bacteriophage T4 polynucleotide kinase. The labeling was performed in a 10 μl reaction at 37°C for 30 min. The enzyme was heat-inactivated at 95°C for 2 min. To remove salts, unincorporated nucleotides and the enzyme, p6AT-1 was gel purified from a 20% polyacrylamide TBE gel with 7M urea. The RNA was extracted from the gel slice and precipitated with 300mM NaOAc and 2.5 volumes 100% ethanol. After a spin at 14,000g and 4°C the RNA pellet was washed twice with chilled 70% ethanol. The pellet was air dried and subsequently resuspended in 50 μl water. 0.5 μl were spotted onto a filter paper disk and counted in a scintillation counter (Cherenkov analysis).

The end-labeling resulted in a p6AT-1 preparation with an activity of 640,000 cpm/μl.

Example 17. Determination of the optimal buffer conditions for SP RNA reactions

E. coli M1 RNA is able to cleave ptRNA^{met} or the minimal substrate p6AT-1 in 100mM NH₄Cl, 100mM TrisCl pH 7.5, 100mM MgCl₂. The conditions for cleavage of either substrate by *S. aureus* SP RNA were unknown and clearly deviated from those optimized for *E. coli*. A large number of different buffer conditions were investigated in order to determine optimal buffer conditions for either substrate cleavage by the *S. aureus* ribozyme.

Cleavage reactions (20 μl) were all performed at 37°C for 30 min with 100nM SP RNA and an additional 200nM SP protein in the holoenzyme reactions. Substrates (ptRNA^{met} or p6AT-1) were added in trace amounts (single-turnover reactions). Cleavage was monitored by size resolution of the intact precursor and the cleaved leader sequence by denaturing gel electrophoresis and autoradiography or visualization on the phosphorimager (Molecular Dynamics Storm 860).

Example 18 Effect of KCl versus NH₄Cl on p6AT-1 cleavage

All experiments were performed in 100mM TrisCl pH7.5, 100mM MgCl₂ and 1M or 2M monovalent salt (KCl or NH₄Cl).

There was hardly any cleavage product detectable at 1M KCl or NH₄Cl. Cleavage occurred when 2M monovalent salt were used. Potassium chloride gave better results than ammonium chloride.

Further optimization of the monovalent salt concentration was investigated following optimization of additional parameters described in the following sections.

Example 19. Effect of the pH on p6AT-1 cleavage

Buffers containing 100mM MgCl₂, 2M KCl and 100mM TrisCl at a pH of 7.0, 7.5 and 8.0 were tested for the ability to promote p6AT-1 cleavage by SP RNA.

The substrate was cleaved in all three buffers but cleavage was very poor at pH 7.0 and optimal at pH 8.0.

Example 20. Influence of PEG 8,000 on p6AT-1 cleavage

The addition of polyethylene glycol has been shown to improve substrate cleavage by *E. coli* M1 RNA. Here the effect of different PEG 8,000 concentrations on p6AT-1 cleavage by SP RNA was demonstrated. 1%, 2.5% and 5% PEG were added to reactions containing 1M KCl, 100mM TrisCl pH8, 100mM MgCl₂.

The ability to cleave the substrate improved with increasing PEG concentrations, and therefore 5% PEG was selected to be included in subsequent reactions.

Example 21. MgCl₂ requirements for p6AT-1 processing

The MgCl₂ concentration was successively reduced from 100mM to 10mM to determine the lowest magnesium ion concentration at which the SP RNA could still process the substrate. 100mM, 50mM, 25mM and 10mM MgCl₂ were tested in a buffer containing 2M KCl, 100mM Tris.Cl pH8 and 5% PEG.

Cleavage results were obtained at 100mM MgCl₂. However SP RNA was still able to process p6AT-1 at 50mM and to some extent at 25mM MgCl₂. No cleavage could be observed at 10mM MgCl₂.

Example 22. Determination of KCl requirements for p6AT-1 cleavage

The KCl concentration was successively increased from 40mM to 1.5M in 100mM Tris.Cl pH8, 100mM MgCl₂ and 5% PEG.

No cleavage occurred at KCl concentrations below 400mM. At 400mM KCl p6AT-1 processing was still very poor but increased with the salt concentration to reach a maximal level of 27% at 1.5M KCl. That was comparable to p6AT-1 cleavage by *E. coli* M1 RNA, where 35% of the substrate was cleaved after 20min. The best buffer conditions for p6AT-1 cleavage by SP RNA were 1.5M KCl, 100mM Tris.Cl pH8, 100mM MgCl₂, 5% PEG.

Example 23. KCl requirements for ptRNA^{met} cleavage by SP RNA

ptRNA^{met} was cleaved under the same buffer conditions as described above. ptRNA^{met} was a better substrate than p6AT-1. By contrast to p6AT-1, ptRNA^{met} could be processed at KCl concentrations below 400mM. Some cleavage could already be detected at 20mM KCl, whereas no cleavage of p6AT-1 was observed under these conditions. The cleavage rate increased with the KCl concentration. At 150mM KCl more than 50% of the substrate was processed, reaching 82% at 800mM KCl. At concentrations of 1M or above 85% of the ptRNA were cleaved, and no further increase could be detected, indicating that 15% of the substrate was uncleavable. *E. coli* M1 RNA as a positive control cleaved 80% of the substrate in 100mM Tris.Cl pH7.5, 100mM NH₄Cl, 100mM MgCl₂ within 20min. The best buffer conditions for ptRNA^{met} cleavage by SP RNA were 1M KCl, 100mM Tris.Cl pH8, 100mM MgCl₂, 5% PEG.

Example 24. Determination of the optimal buffer conditions for *S. aureus* RNase P holoenzyme reactions: p6AT-1 cleavage by the holoenzyme

The *S. aureus* RNase P holoenzyme cleavage of p6AT-1 was investigated over a range of different KCl concentrations at low magnesium ion concentrations. All reactions were performed in 100mM Tris.Cl pH8, 10mM MgCl₂, 5% PEG and 40mM to 1.5M KCl. Cleavage could already be observed at 40mM KCl. The cleavage rate increased with the concentration of monovalent salt, peaking at 150mM, where 50% of the substrate was processed. At higher KCl concentrations the cleavage rate decreased again and p6AT-1 processing did not occur at KCl concentrations of 600mM and above. The *E. coli* holoenzyme control cleaved 57% of the p6AT-1 within 20min in 100mM Tris.Cl pH7.5, 100mM NH₄Cl, 10mM MgCl₂. The optimal buffer for p6AT-1 cleavage by the holoenzyme was 150mM KCl, 100mM Tris.Cl pH 8.0, 10mM MgCl₂, 5% PEG 8000. p6AT-1 was a better substrate for the holoenzyme than for SP RNA, as under optimal conditions only 27% p6AT-1 was cleaved as opposed to 50% by the *S. aureus* RNase P holoenzyme.

Example 25. MgCl₂ requirements for ptRNA cleavage by the holoenzyme

The *S. aureus* RNase P holoenzyme failed to cleave ptRNA^{met} under the same conditions employed for p6AT-1 processing. The holoenzyme was not able to cleave the substrate at 10mM MgCl₂ under any given KCl concentration. The concentration of magnesium ions was increased to 20mM and ptRNA cleavage by the holoenzyme as well as the SP RNA ribozyme were tested at different KCl concentrations. The buffers contained 100mM Tris.Cl pH8, 20mM MgCl₂, 5% PEG and 10mM, 50mM, 150mM, 600mM or 1.5M KCl.

Cleavage of ptRNA by the holoenzyme at 10mM MgCl₂ and low KCl concentrations was very poor. At 20mM MgCl₂ and 150mM KCl the holoenzyme processed ptRNA very well, 54% of the substrate was cleaved within a 20min incubation. Only a low percentage of substrate processing could be detected at 600mM KCl, but good cleavage occurred at 1.5M KCl. The SP RNA alone was also able to cleave ptRNA under these conditions but was unable to cleave ptRNA at lower salt concentrations. Therefore the optimal MgCl₂ concentration for ptRNA^{met} cleavage by the holoenzyme was 20mM.

Example 26 Determination of the optimal KCl concentration

The *S. aureus* RNase P holoenzyme was tested for its ability to cleave ptRNA^{met} under different KCl concentrations

ranging from 20mM to 1.5M. The buffers contained 100mM Tris-Cl pH8, 20mM MgCl₂, 5% PEG 8000 and xM KCl.

The holoenzyme was able to cleave ptRNA^{met} at KCl concentrations as low as 20mM. The cleavage rate increased with the salt concentration with a peak at 150mM where 55% of the substrate was processed. The cleavage rate decreased at KCl concentrations above 200mM but improved again at 1M KCl reaching a maximum at 1.5M. The cleavage at high salt concentrations was not due to holoenzyme activity but result of SP RNA ribozyme activity at high KCl concentrations.

ptRNA^{met} was a better substrate for SP RNA than for the *S. aureus* RNase P holoenzyme. The optimal buffer conditions for ptRNA^{met} cleavage by the holoenzyme were 150mM KCl, 100mM Tris Cl pH8, 20mM MgCl₂, 5% PEG.

Example 26 Strain selection, Library Production and Sequencing

The polynucleotide having the DNA sequence given in SEQ ID NO: 1 was obtained from a library of clones of chromosomal DNA of *Staphylococcus aureus* in *E. coli*. The sequencing data from two or more clones containing overlapping *Staphylococcus aureus* DNAs was used to construct the contiguous DNA sequence in SEQ ID NO: 1. Libraries may be prepared by routine methods, for example:

Methods 1 and 2 below.

Total cellular DNA is isolated from *Staphylococcus aureus* WCUH 29 according to standard procedures and size-fractionated by either of two methods.

Method 1

Total cellular DNA is mechanically sheared by passage through a needle in order to size-fractionate according to standard procedures. DNA fragments of up to 1 kbp in size are rendered blunt by treatment with exonuclease and DNA polymerase, and EcoRI linkers added. Fragments are ligated into the vector Lambda ZapII that has been cut with EcoRI, the library packaged by standard procedures and *E. coli* infected with the packaged library. The library is amplified by standard procedures.

Method 2

Total cellular DNA is partially hydrolyzed with a one or a combination of restriction enzymes appropriate to generate a series of fragments for cloning into library vectors (e.g., RsaI, PstI, AluI, Bsh1235I), and such fragments are size-fractionated according to standard procedures. EcoRI linkers are ligated to the DNA and the fragments then ligated into the vector Lambda ZapII that have been cut with EcoRI, the library packaged by standard procedures, and *E. coli* infected with the packaged library. The library

SEQUENCE LISTING

5

(1) GENERAL INFORMATION

10

(i) APPLICANT: Guth, Sabine
Jennings, Joanne
Prescott, Catherine

15

(ii) TITLE OF THE INVENTION: NOVEL RNASE P

20

(iii) NUMBER OF SEQUENCES: 14

(iv) CORRESPONDENCE ADDRESS:

25

(A) ADDRESSEE: SmithKline Beecham Corporation
(B) STREET: 709 Swedeland Road
(C) CITY: King of Prussia
(D) STATE: PA
(E) COUNTRY: USA
(F) ZIP: 19406-0939

30

(v) COMPUTER READABLE FORM:

35

(A) MEDIUM TYPE: Diskette
(B) COMPUTER IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

40

(vi) CURRENT APPLICATION DATA:

45

(A) APPLICATION NUMBER
(B) FILING DATE: 06-JUN-1997
(C) CLASSIFICATION

50

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER 60/019,234
(B) FILING DATE: 06-JUN-1996

55

(A) APPLICATION NUMBER 60/029,928

(B) FILING DATE: 01-NOV-1996

(A) APPLICATION NUMBER 60/029,928

(B) FILING DATE: 01-NOV-1996

(viii) ATTORNEY/AGENT INFORMATION

(A) NAME: Gimmi, Edward R

(B) REGISTRATION NUMBER 38,891

(C) REFERENCE/DOCKET NUMBER P50493

(ix) TELECOMMUNICATION INFORMATION

(A) TELEPHONE: 610-270-4478

(B) TELEFAX: 610-270-5090

(C) TELEX:

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 354 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ATGTTATTGG AAAAAGCTTA CCGAATTAAA AAGAATGCAG ATTTTCAGAG AATATATAAA	60
AAAGGTCATT CTGTAGCCAA CAGACAATTT GTTGTATACA CTTGTAATAA TAAAGAAATA	120
GACCATTTTC GCTTAGGTAT TAGTGTTCCT AAAAACTAG GTAATGCAGT GTTAAGAAAC	180
AAGATTAAAA GAGCAATACG TGAAAATTTT AAAGTACATA AGTCGCATAT ATTGGCCCAA	240
GATATTATTG TAATAGCAAG ACAGCCAGCT AAAGATATGA CGACTTTACA AATACAGAA	300
AGTCTTGAGC ACGTACTTAA AATTGCCCAA GTTTTAAATA AAAAGATTAA GTAA	354

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 117 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

Met Leu Leu Glu Lys Val Tyr Arg Ile Lys Lys Asn Ala Asp Phe Gly
 1           5           10           15
Arg Ile Tyr Lys Lys Gly His Ser Val Ala Asn Arg Gln Phe Val Val
          20           25           30
Tyr Thr Cys Asn Asn Lys Glu Ile Asp His Phe Arg Leu Gly Ile Ser
25           35           40           45
Val Ser Lys Lys Leu Gly Asn Ala Val Leu Arg Asn Lys Ile Lys Arg
 50           55           60
Ala Ile Arg Glu Asn Phe Lys Val His Lys Ser His Ile Leu Ala Lys
30           65           70           75           80
Asp Ile Ile Val Ile Ala Arg Gln Pro Ala Lys Asp Met Thr Thr Leu
          85           90           95
Gln Ile Gln Asn Ser Leu Glu His Val Leu Lys Ile Ala Lys Val Phe
          100          105          110
Asn Lys Lys Ile Lys
          115

```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 401 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

5	GTTCTGATAT TTGGGTAAT CGCTATATTA TATAGAGGAA AGTCCATGCT CCACAGTCTG	60
	AGATGCATTG TAGTGTTTGT GCTTGATGAA ACAATAAATC AAGGCATTAA TTTGACGGCA	120
	ATGAAATATC CTAAGTCTTT CGATATGGAT AGAGTAATTT GAAAGTGCCA CAGTGACGTA	180
10	GCTTTTATAG AAATATAAAA GGTGGAACGC GGTAAACCCC TCGAGTGAGC AATCCAAATT	240
	TGGTAGGAGC ACTTGTTTAA CGGAATTCAG CGTATAAAGC AGACACACTT CGCGAAATGA	300
	AGTGGTGTAC GACAGATGGT TATCACCTGA GTACCAAGTGT GACTAGTGCA CGTGATGAGT	360
15	ACGATGGAAC AGAACATGGC TTATAGAAAT ATCACTACTA G	401

(2) INFORMATION FOR SEQ ID NO: 4:

20 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 42 base pairs

(B) TYPE: nucleic acid

25 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: RNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

35	GAAUUCGCCC GGACUCGGUU CAGUUCGAG UCGGGCACC AC	42
----	--	----

(2) INFORMATION FOR SEQ ID NO: 5:

40 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

45 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

55

CGCGAAGTGT GTCTCGTTTA TACG

24

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GAGGAAAGTC CATGCTC

17

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

CGTATTGCTC TTTTAATCTT GTTT

24

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 85 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GTTCCTACCTAT TTGCAAACAG ACTCCATTCT GAAGCCCTAC GCAACTCCCC 60
AATTACAAA GTAAGGGATG GGATA 85

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 213 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GTTTT CATT C	CCTACCTAT	CCTTACTTAA	TCTTTTATT	AAAAAGAATG	CAGATTTTCA	60
GAGAATATAT	AAAAAAGGT C	ATTCTGTAGC	CAACAGACAA	TTTGTTGTAT	ACACTTGTAA	120
TAA TAAAGAA	ATAGACCATT	TT CGCTTAGG	TATTAGTGTT	TCTAAAAAAC	TAGGTAATGC	180
AGTGTTAAGA	ACAAGATTAA	AAGAGCAATA	CGT			213

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 215 base pairs

(3) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

ACGATAAATA ACACTATGTT ATTGGAAAAA GTCTACCGAA TTAAAAAGAA TGCAGATTTT 60
 CAGAGAATAT ATAAAAAAGG TCATTCTGTA GCCAACAGAC AATTGTTGT ATACACTTGT 120
 AATAATAAAG AAATAGACCA TTTTCGCTTA GGTATTAGTG TTTCTAAAAA ACTAGGTAAT 180
 GCAGTGTTAA GAACAAGATT AAAAGAGCAA TACGT 215

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 170 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GAAAATTTCA AAGTACATAA GTCGCATATA TTGGCCAAAG ATTGTAATAG CAAGACAGCC 60
 AGCTAAAGAT ATGACGACTT TACAAATACA GAATAGTCTT GAGCAGTAC TTAAAATTGC 120
 CAAAGTTTTT AATAAAAAGA TTAAGTAAGG ATAGGTTAGG GAATGAAAAC 170

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 93 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GGGCGAAUUC GCCUCGGCUA CGUAGCUCAG UUGGUUAGAG CACAUCACUC AUAUGAUGG 60
 GGUACAGGU UCGAAUCCCG UCGUAGCCAC CAG 93

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

YGATATTTCK GRTAAYC

17

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 137 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TGNAGACAGA TGGTTATCAC CTGAGTACCA GTGTGACTAG TGCACGTGAT GAGTACGATG	60
GAACAGAACA TGGCTTATAG AAATATCACT ACTAGTTTAG CTCTCCTAGA TGATGGAGAG	120
CTTTTTTTCAT GAAAAGA	137

Claims

1. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

(a) a polynucleotide having at least a 50% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of SEQ ID NO:2;

(b) a polynucleotide having at least a 50% identity to a polynucleotide comprising the nucleic acid sequence of SEQ ID NO:1, 3 or 4;

(c) a polynucleotide having at least a 50% identity to a polynucleotide encoding the same mature polypeptide expressed by the RNaseP gene contained in the *Staphylococcus aureus* of the deposited strain;

(d) a polynucleotide having at least a 50% identity to a polynucleotide comprising the same catalytic RNA expressed by the RNaseP RNA gene contained in the *Staphylococcus aureus* of the deposited strain;

(e) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 70% identical

to the amino acid sequence of SEQ ID NO:2;

(f) a polynucleotide which is complementary to the polynucleotide of (a), (b), (c), (d) or (e); and

(g) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c), (d), (e) or (f).

- 5 2. The polynucleotide of Claim 1 wherein the polynucleotide is DNA.
3. The polynucleotide of Claim 1 wherein the polynucleotide is RNA.
4. The polynucleotide of Claim 2 comprising the nucleic acid sequence set forth in SEQ ID NO:1.
- 10 5. The polynucleotide of Claim 2 comprising nucleotide 1 to 354 set forth in SEQ ID NO:1.
6. The polynucleotide of Claim 2 which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2.
- 15 7. A vector comprising the polynucleotide of Claim 1.
8. A host cell comprising the vector of Claim 7.
9. A process for producing a polypeptide comprising: expressing from the host cell of Claim 8 a polypeptide encoded by said DNA.
- 20 10. A process for producing a RNaseP polypeptide or fragment comprising culturing a host of claim 8 under conditions sufficient for the production of said polypeptide or fragment.
- 25 11. A polypeptide comprising an amino acid sequence which is at least 50% identical to the amino acid sequence of SEQ ID NO:2.
12. A polypeptide comprising an amino acid sequence as set forth in SEQ ID NO:2.
- 30 13. An antibody against the polypeptide of claim 11.
14. An antagonist which inhibits the activity or expression of the polypeptide of claim 11.
- 35 15. A method for the treatment of an individual in need of RNaseP polypeptide comprising: administering to the individual a therapeutically effective amount of the polypeptide of claim 11.
16. A method for the treatment of an individual having need to inhibit RNaseP polypeptide comprising: administering to the individual a therapeutically effective amount of the antagonist of Claim 14.
- 40 17. A process for diagnosing a disease related to expression or activity of the polypeptide of claim 11 in an individual comprising:
 - (a) determining a nucleic acid sequence encoding said polypeptide, and/or
 - (b) analyzing for the presence or amount of said polypeptide in a sample derived from the individual.
- 45 18. A method for identifying compounds which interact with and inhibit or activate an activity of the polypeptide of claim 11 comprising:
 - 50 contacting a composition comprising the polypeptide with the compound to be screened under conditions to permit interaction between the compound and the polypeptide to assess the interaction of a compound, such interaction being associated with a second component capable of providing a detectable signal in response to the interaction of the polypeptide with the compound;
 - and determining whether the compound interacts with and activates or inhibits an activity of the polypeptide by detecting the presence or absence of a signal generated from the interaction of the compound with the polypeptide.
- 55 19. A method for inducing an immunological response in a mammal which comprises inoculating the mammal with RNaseP polypeptide of claim 11, or a fragment or variant thereof, adequate to produce antibody and/or T cell

immune response to protect said animal from disease.

20. A method of inducing immunological response in a mammal which comprises delivering a nucleic acid vector to direct expression of RNaseP polypeptide of claim 11, or fragment or a variant thereof, for expressing said RNaseP polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response to produce antibody and/ or T cell immune response to protect said animal from disease.
21. An isolated RNase P from *Staphylococcus aureus*.
22. The RNase P according to Claim 1 having an RNA component of the sequence of SEQ ID NO: 1.
23. The RNase P according to Claim 1 comprising a protein component of the sequence of SEQ ID NO: 2.
24. An isolated DNA encoding the components of Claim 23.
25. A method for individual or co-expression of the DNA of Claim 24.
26. A screen for the identification of inhibitors of RNase P comprising one or more of the RNase P components set forth in SEQ ID NO: 1 through 10.
27. An isolated RNA component of Staphylococcus aureus RNase P.
28. An isolated protein component of Staphylococcus aureus RNase P.
29. An isolated DNA encoding the components of Claim 22.
30. A polynucleotide selected from the group consisting of SEQ ID NO: 5 through 12.

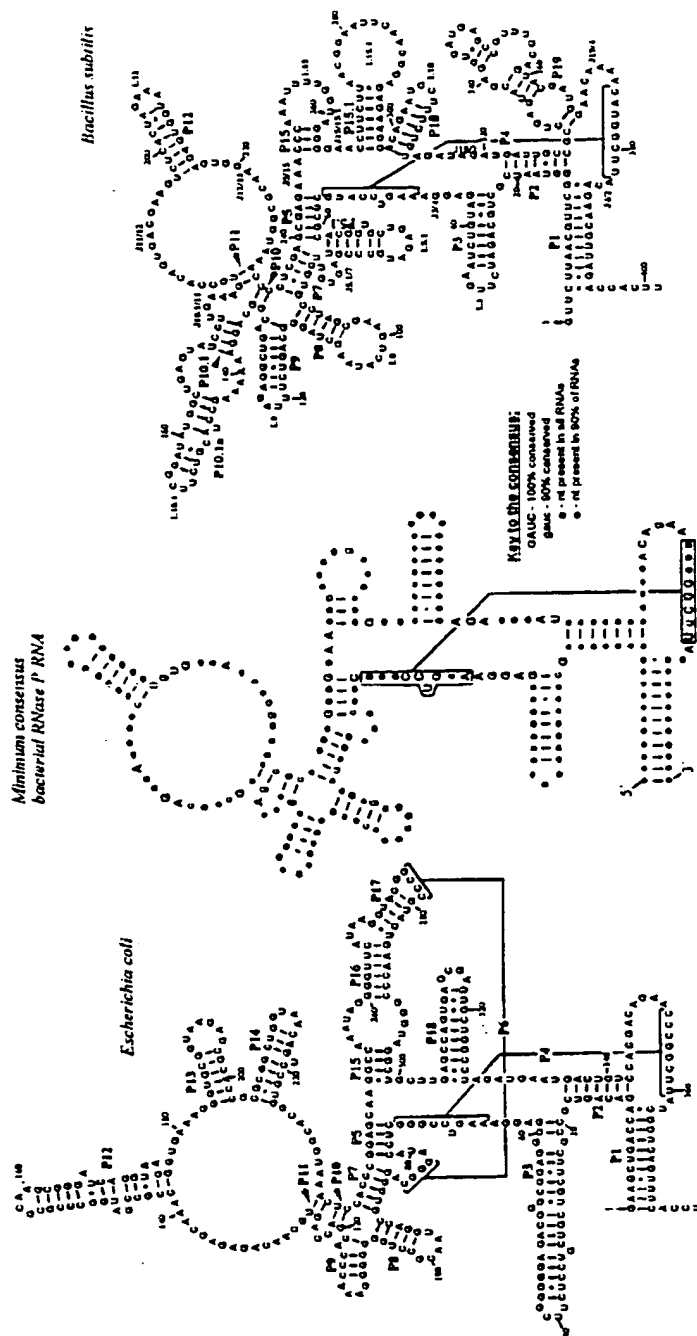


Figure 1.

Figure 2.

	10	20	30	40	50	60	70
Eco	MVKLAFFPRELPLLT	PSQFTFVFQQP	QORAGTPOITIL	GRNLNSLG-----	HPRIGLTVAKKNV	RRRAHE	
Pmi	MVKLAFFPRELRLLT	PKHFNFVFQQP	QRASSPEVTIL	GRQNELG-----	HPRIGLTIAKKNV	KRAHE	
Sbi	----MLPTENRLRR	REDFATAVRRG	RGRPLLVHRL	SGATDPHAPGES	APPTRAGFVVS-	KAVGGAVV	
Mlu	----MLPRDRRVRT	PAEFRHLGRTG	TRAGRRTVVV	SVATDPDQTR	STSPSAPRPRAG	FVVS-KAVGNAVT	
Bsu	--MSHLKKRNRL	KKNEFDQKVF	KHGTSVANRQ	FVL----YTLDQ	PENDE----LRVGL	SVSKK-IGNAVM	
Sau	--ML-LEKAYRI	KNADFQRIYK	KGHSVANRQ	FW----YTCNN	KEIDH----FRLGI	SVSKK-LGNAVL	
	80	90	100	110	120	130	
Eco	RNRKRLTRESFRL--	ROHELPAMD	FVVVAKK--	GVADLDNRAL	SEALEKLWRR	HCR--LARGS---	
Pmi	RNRKRLAREYFRL--	HQHQLPAMD	FVVLVRK--	GVAELDNHQL	TEVLGKLWRR	HCR--LAQKS---	
Sbi	RNQVKRRRLRHL	VCD--RLSAL	PPGSLVV-V	RALPGAGDAD	HAQLARDLDA	ALQR----LLGGG	TR--
Mlu	RNRVKRRRLRAV	VAEQMRLP	PLRDLPLVQ	VRALPAAAE	ADYALLRRET	VGALGKALKP	HLPAASEHA
Bsu	RNRKRLIRQAFLE-	EKER-LKE	KDYIIIAK--	PASOLTYEET	KKSLQHLFR	KSSL--YKKSSSK-	
Sau	RNKIKRAIREN	FKV-HKSHILA-	KDIIVIA	RQ--PAKDM	TTLQIONS	LEHVLKIAK	V--FNKKIK

Escherichia coli (Eco), *Proteus mirabilis* (Pmi), *Streptomyces bikiniensis* (Sbi),
Micrococcus luteus (Mlu), *Bacillus subtilis* (Bsu), *Staphylococcus aureus* (Sau).

Figure 4.

Bacillus subtilis

S. aureus

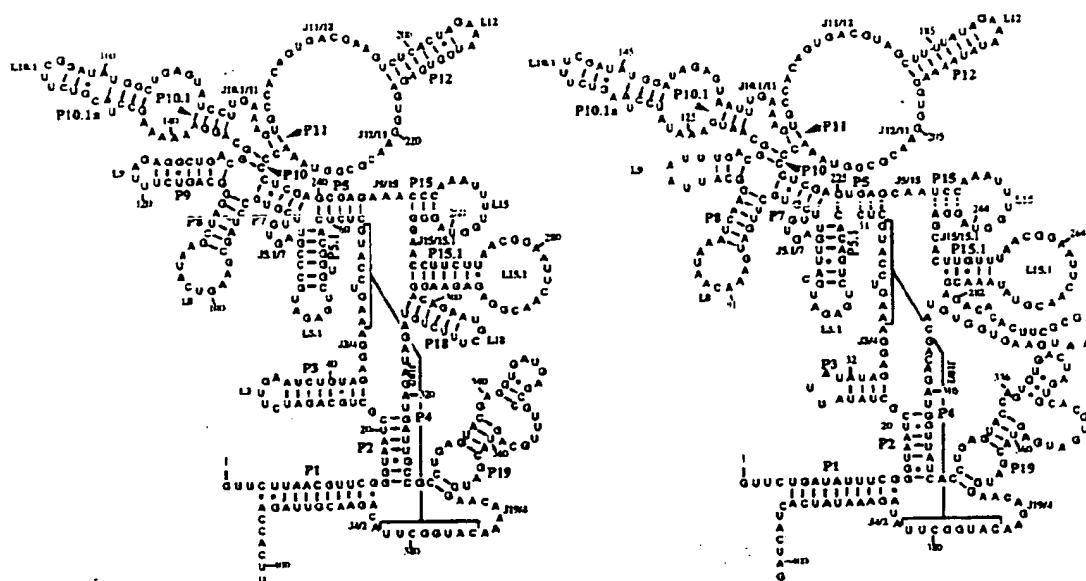


Figure 5.

```

ATG TTA TTG GAA AAA GCT TAC CGA ATT AAA AAG AAT 36
M   L   L   E   K   A   Y   R   I   K   K   N
GCA GAT TTT CAG AGA ATA TAT AAA AAA GGT CAT TCT 72
A   D   F   Q   R   I   Y   K   K   G   H   S
GTA GCC AAC AGA CAA TTT GTT GTA TAC ACT TGT AAT 108
V   A   N   R   Q   F   V   V   Y   T   C   N
AAT AAA GAA ATA GAC CAT TTT CGC TTA GGT ATT AGT 144
N   K   E   I   D   H   F   R   L   G   I   S
GTT TCT AAA AAA CTA GGT AAT GCA GTG TTA AGA AAC 180
V   S   K   K   L   G   N   A   V   L   R   N
AAG ATT AAA AGA GCA ATA CGT GAA AAT TTC AAA GTA 216
K   I   K   R   A   I   R   E   N   F   K   V
CAT AAG TCG CAT ATA TTG GCC AAA GAT ATT ATT GTA 252
H   K   S   H   I   L   A   K   D   I   I   V
ATA GCA AGA CAG CCA GCT AAA GAT ATG ACG ACT TTA 288
I   A   R   Q   P   A   K   D   M   T   T   L
CAA ATA CAG AAT AGT CTT GAG CAC GTA CTT AAA ATT 324
Q   I   Q   N   S   L   E   H   V   L   K   I
GCC AAA GTT TTT AAT AAA AAG ATT AAG TAA 354
A   K   V   F   N   K   K   I   K   stop

```

Figure 6.

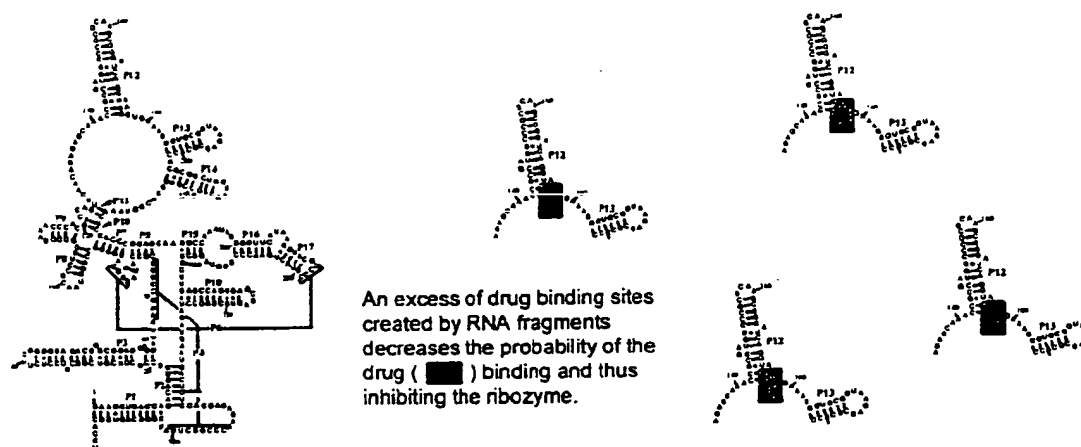


Figure 7.

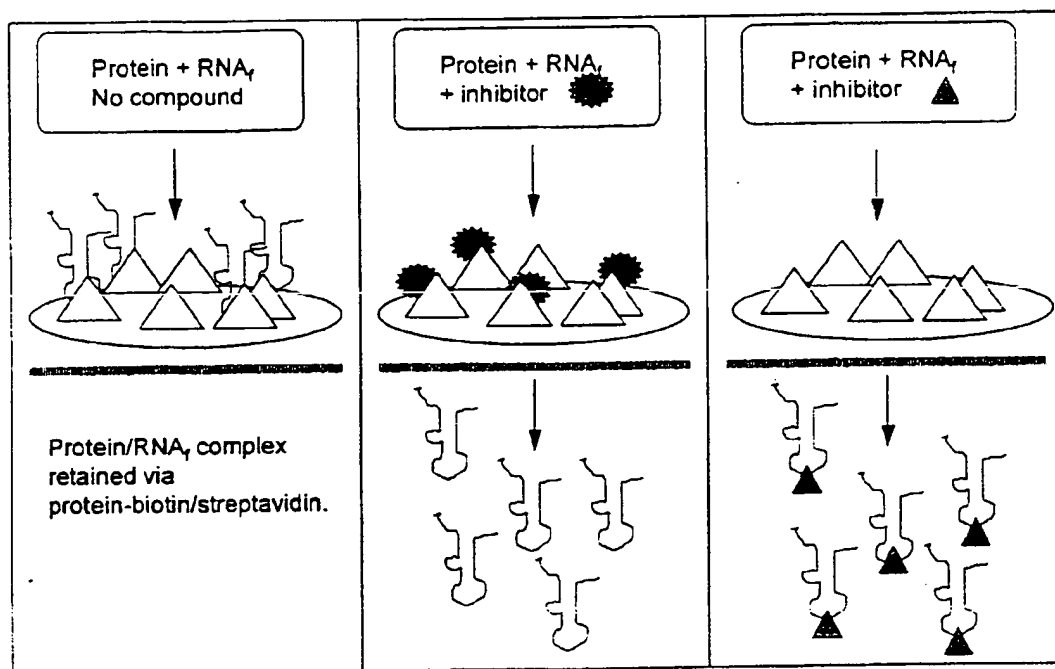


Figure 8.

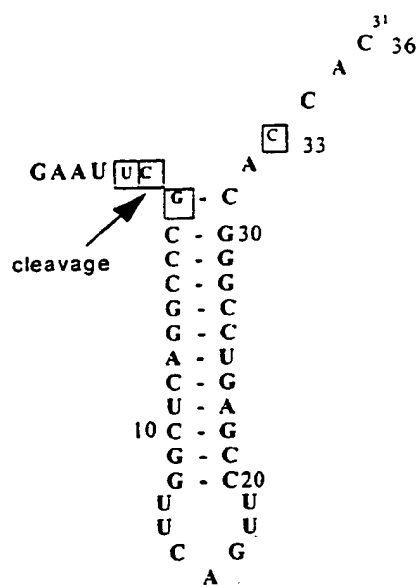


Figure 9.

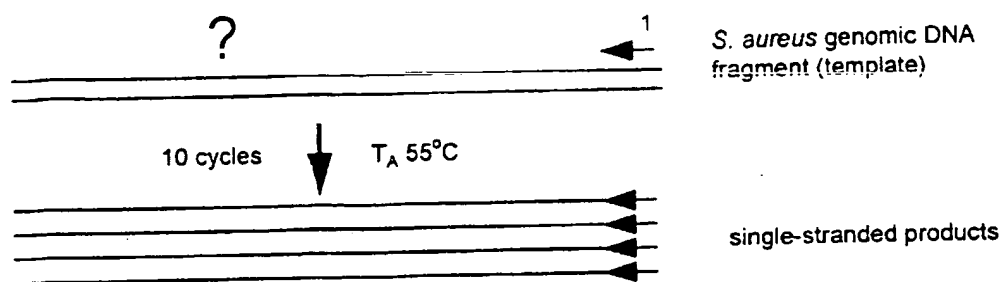


Figure 10.

